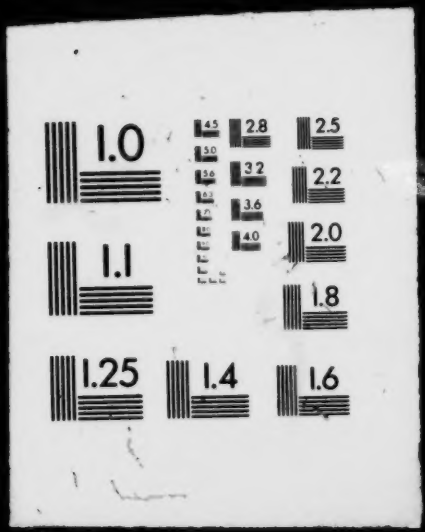
1 OF 1
PB 89
211726



PB89-211726

Alternatives to Conventional Risk Assessment in Determining Appropriate Cleanup Levels for Superfund Remediation (Question 3105)

North Carolina Univ. at Chapel Hill

Prepared for:

Environmental Protection Agency, Washington, DC

1988

U.S. Department of Commerce
National Technical Information Service
NTIS

Final
3105
PB89-211726

Alternatives to Conventional Risk Assessment in Determining Appropriate Cleanup Levels for Superfund Remediation

(Question 3105)

Prepared for the
EPA's National Network for
Environmental Policy Studies

ROBERT G. HETES

Department of Environmental Sciences and Engineering
School of Public Health
University of North Carolina at Chapel Hill

REPRODUCED BY
U.S. DEPARTMENT OF COMMERCE
NATIONAL TECHNICAL INFORMATION SERVICE
SPRINGFIELD, VA. 22161

50272 - 101

REPORT DOCUMENTATION PAGE	1.REPORT NO. 600/9-89/066	2.	3. Recipient's Accession PB89 211726/AS

4. Title and Subtitle
Alternatives to Conventional Risk Assessment in Determining Appropriate Cleanup Levels for Superfund Remediation

5. Report Date
Completed Fall 1988

6.

7. Author(s)
Robert G. Hetes, Graduate Student

8. Performing Organization Rept. No.

9. Performing Organization Name and Address
Department of Environmental Sciences and Engineering
School of Public Health
University of North Carolina
Chapel Hill, North Carolina 27599

10. Project/Task/Work Unit No.

11. Contract(C) or Grant(G) No.
(C)
(G) U-913-012-01-0

12. Sponsoring Organization Name and Address
U. S. Environmental Protection Agency
Office of Cooperative Environmental Management
401 M Street St., SW A101-F6
Washington, DC 20460

13. Type of Report & Period Covered
Technical Report

14.

15. Supplementary Notes

16. Abstract (Limit: 200 words)

This report is part of the National Network for Environmental Management Studies conducted under the auspices of the Office of Cooperative Environmental Management - U. S. Environmental Protection Agency (EPA). The report examines the question "Do alternatives exist to the use of conventional risk assessment in establishing clea up levels at CERCLA remedial actions?" The Comprehensive Environmental Response, Compensation and Liability Act of 1980 (CERCLA), more commonly called Superfund, established a national program for responding to releases of hazardous substances into the environment. Central to the Superfund program are the questions of which sites are deserving of remedial action and to what extent they should be cleaned. The first is answered by the use of the National Priorities List (NPL) which ranks the sites according to a hazard assessment. The second, the topic of this paper, is not as well defined.

This report includes a literature review followed by a series of personal interviews with individuals (in both the public and private sectors) with interest in CERCLA activities as well as a summary and analysis of the existing method of remedial action selection, both in design and practice. The consensus is that the existing process as designed is adequate, but problems arise from inconsistent application. The Conventional risk assessment is found to be limited in addressing multi-compound and muti-media situations. The report identifies four alternatives to the conventional risk assessment.

Conclusions drawn in this report include: 1) concentration-based standards yielding consistent outcomes inadequately address additivity, exposure, costs, and site characteristics; 2) best demonstrated available technology (BDAT) holds the greatest future promise for acceptability and effectiveness but is presently inadequate due to limited existing experience; 3) benefit-cost analysis is not viable because CECLA issues are not conducive to economic quantification; 4) improved risk assessment has the greatest short-term potential — augmenting existing methods, not requiring extensive change, and better addressing multiple contaminant and media situations.

17. Document Analysis a. Document Descriptors

b. Identifier's/Open-Ended Terms

c. COSATI Field/Group

18. Availability Statement	19. Security Class (This Report) Unclassified	21. No. of Pages 96
	20. Security Class (This Page) Unclassified	22. Price

(See ANSI-Z39.18)

OPTIMAL FORM 272 (4-77)
(Formerly NTIS-35)

Alternatives to Conventional Risk Assessment in Determining Appropriate Cleanup Levels for Superfund Remediation

(Question 3105)

Prepared for the
EPA's National Network for
Environmental Policy Studies

ROBERT G. HETES

Department of Environmental Sciences and Engineering
School of Public Health
University of North Carolina at Chapel Hill

Disclaimer

This report was furnished to the U. S. Environmental Protection Agency by the graduate student identified on the cover page, under a National Network of Environmental Policy Studies fellowship.

The contents are essentially as received from the author. The opinions, findings, and conclusions expressed are those of the author and not necessarily those of the Environmental Protection Agency. Mention, if any, of company, process, or product names is not to be considered as an endorsement by the U. S. Environmental Protection Agency.

TABLE OF CONTENTS

Prologue

A. Introduction. 1

B. Summary of Remedial Action Selection Process. . . . 4

C. Role of Risk Assessment in Remedial Action Selection. 10

D. Critique of Existing Remedial Action Action Selection Process. 21

E. Critique of Existing Risk Assessment Methodology. . 27

F. Alternatives. 36

I. Absolute: Uniform Concentration Based Stds. . 37

II. Relative: (a) Designated Treatment Technology . 43

(b) Improved Risk Assessment 49

(c) Benefit-Cost Analysis. 58

G. Evaluation of Alternatives. 66

H. Future Study Needs. 76

I. Conclusions and Recommendations 79

J. Appendix. 83

K. Bibliography. . . . . .

PROLOGUE

The research discussed in this paper was supported by the EPAs 1988 National Network of Environmental Policy Studies program. The work was sponsored by the Office of Emergency and Remedial Response within the Office of Solid Waste and Emergency Response, and conducted under the supervision of the Toxics Integration Branch and the Policy and Analysis Staff.

The initial phase of the study required a review of existing EPA documents and guidances regarding the selection of remedial actions and establishing cleanup levels. In addition to internal EPA documents, reports or criticisms of the Superfund program by outside agencies and organizations were also reviewed.

Once all written reports were reviewed, a series of personal interviews was conducted with personnel involved in the preparation of the above documents and in day to day management of the Superfund program or related activities. Persons were interviewed either in person or by telephone and represented both the EPA and outside interests such as public interest groups, industrial organizations and other private interests. These interviews were conducted in confidence and the interviewees were assured of anonymity. Such confidentiality, while yielding the greatest level of honesty and criticism, does not allow for specific references or citations to specific information and statements within the report. For this reason certain claims will be referred to but not specifically referenced.

To aid in the reference to claims in the report the following citation notation will be used:

(Category 1) refers to individuals within the EPA
(Category 2) refers to governmental officials not associated with the EPA
(Category 3) refers to individuals within public interests groups (e.g., environmental groups)
(Category 4) refers to individuals within private interest organizations

A list of documents reviewed can be found in the bibliography and a list of interviewees and their affiliation can be found in the appendix.

Any analytical problem can be interpreted in different manners. For this paper I will use four major criteria of analysis; completeness of conception, rigor of thought, consistency of application, and demonstrability of goals. Completeness of conception refers to the effort of defining and including in the analysis all influencing parameters, or the breadth of the analysis. The rigor of thought refers to

the level of detail afforded any particular parameter, the extent that evidence can be related to judgement, or the depth of analysis. Rigor is met through the development of formal procedures representing that parameter. The consistency of application is failrly self explanatory and refers to how the analysis or use of a parameter is used throughout the analysis. The demonstrability of goals refers to the ability of any proposal to meet or achieve the goals and whether those goals are explicitly carried out.

The four elements discussed above are in continual opposition to one another. To concentrate on or improve one of the four elements detracts from the others. For example, the greatest amount of rigor yields the least amount of completeness of conception. The greatest rigor would occur in a sampling problem and would rely entirely on statistics or empirical evidence. However, this problem is not capable of incorporating theoretical or subjective judgements, and therefore does not meet the goals of completeness of conception.

To critique any model or process one must keep these facts in mind, that tradeoffs between the four elements are unavoidable. To develop a alternative model or process the four elements must be balanced into a pragmatic and useable form which achieves the overall goals as best possible. Certain elements may have to be sacrificed to achieve a workable and implementable form.

A. INTRODUCTION

The Comprehensive Environmental Response, Compensation and Liability Act of 1980 (CERCLA), more commonly called Superfund, established a national program for responding to releases of hazardous substances into the environment. In addition, the National Oil and Hazardous Substances Pollution Contingency Plan (NCP) and the Superfund Amendments and Reauthorization Act of 1986 (SARA) further define the goals and objectives of the program and the process for determining appropriate remedial actions at Superfund sites.

Superfund focuses on sites for which those responsible for the past disposal practices, potentially responsible parties (PRP), cannot be identified or no longer exist. Superfund supplies the funds and resources to oversee and implement cleanup at these sites. Central to the Superfund program are the questions of which sites are deserving of remedial action and to what extent they should be cleaned. The first is answered by the use of the National Priorities List (NPL) which ranks the sites according to a hazard assessment. The second, the topic of this paper, is not as well defined.

The extent of cleanup can be thought of as a question of "how clean is clean?" There are essentially two types of results under these remedial actions: clean closure and those with residual contamination. Clean closures are such that the remedial action has removed or treated the contamination to what are considered safe levels with negligible residual risk. However, typically clean closures are infeasible due to expense and/or technical constraints. In these circumstances, the question arises to what level of residual contamination (and concommitant risks) should cleanup occur.

Selection of remedial actions occurs through the Remedial Investigation and Feasibility Study (RI/FS) process. The above laws and associated regulations require that remedial actions chosen be cost-effective and adequate to protect public health and the environment. These cost-effective remedies are specifically required by regulation to be based on a comparison of alternatives that examines public health impacts, environmental impacts, technological and engineering feasibility, cost and institutional factors. Section 121 of SARA also requires that remedies attain or exceed the of applicable or relevant and appropriate requirements (ARAR) under federal or state environmental laws. Statutory requirements are presented in Table A-1.

Table A-1

Specific statutory requirements under CERCLA for the selection of remedial actions are as follows:

be protective of human health and the environment;

attain or exceed applicable or relevant and appropriate requirements of federal or state environmental laws;

be cost-effective;

use permanent solutions and alternative treatment technologies or resource recovery systems to the maximum extent practicable;

use actions that reduce the toxicity, mobility or volume of the waste as a principal element.

In addition to the statutory requirements outlined in Table A-1, CERCLA places an emphasis on considering the long-term effectiveness and related program concerns. To incorporate these concerns regulations require that the elements outlined in Table A-2 should be considered. While these

Table A-2

Statutory Considerations Specified by Regulations (NCP, 1988)

the long-term uncertainties associated with land disposal;

the goals, objectives and requirements of the Solid Waste Disposal Act;

the persistence, toxicity, and mobility of hazardous substances and their constituents, and their propensity to bioaccumulate;

short- and long-term potential for adverse health effects from human exposure;

long-term maintenance costs;

the potential for future remedial actions if the alternative remedial action in question fails; and

the potential threat to human health and the environment associated with excavation, transportation, redisposal, or containment.

considerations have no specific statutory requirements or standards, regulations require they be considered and addressed in any remedial action selection.(NCP, 1988)

To incorporate the above statutory requirements and considerations, the Assistant Administator of the Office of Solid Waste and Emergency Response developed and distributed nine evaluation criteria. These criteria consolidate all of the above concerns as well as additional technical and policy considerations and will be discussed in the revised NCP (1988). These evaluation criteria serve as the basis for conducting the detailed analysis of alternatives required in the Feasibility Study (FS) stage and in subsequent selection of specific remedial actions for cleanup. Additionally any proposed alternatives to the process of remediation selection or the establishment of cleanup levels would have to be appraised on the basis of these evaluation criteria. These evaluation criteria are: (1)short-term effectiveness; (2)long-term effectiveness and permanence; (3)reduction of toxicity, mobility or volume; (4)implementability; (5)cost; (6)compliance with Applicable or Relevant and Appropriate Requirements (ARARs); (7)overall protection of human health and the environment; (8)state acceptance and (9)community acceptance. These evaluation criteria are presented in greater detail in Table B-1.

B. SUMMARY OF THE REMEDIAL ACTION SELECTION PROCESS

The selection of remedial actions is an interactive process incorporating many complex variables, disciplines and considerations. Factors which may differ greatly among sites and influence the final decision include the nature and extent of the contamination, site characteristics, technology available (for both chemical- and site-specific characteristics), costs, ARARs and potentially exposed population.

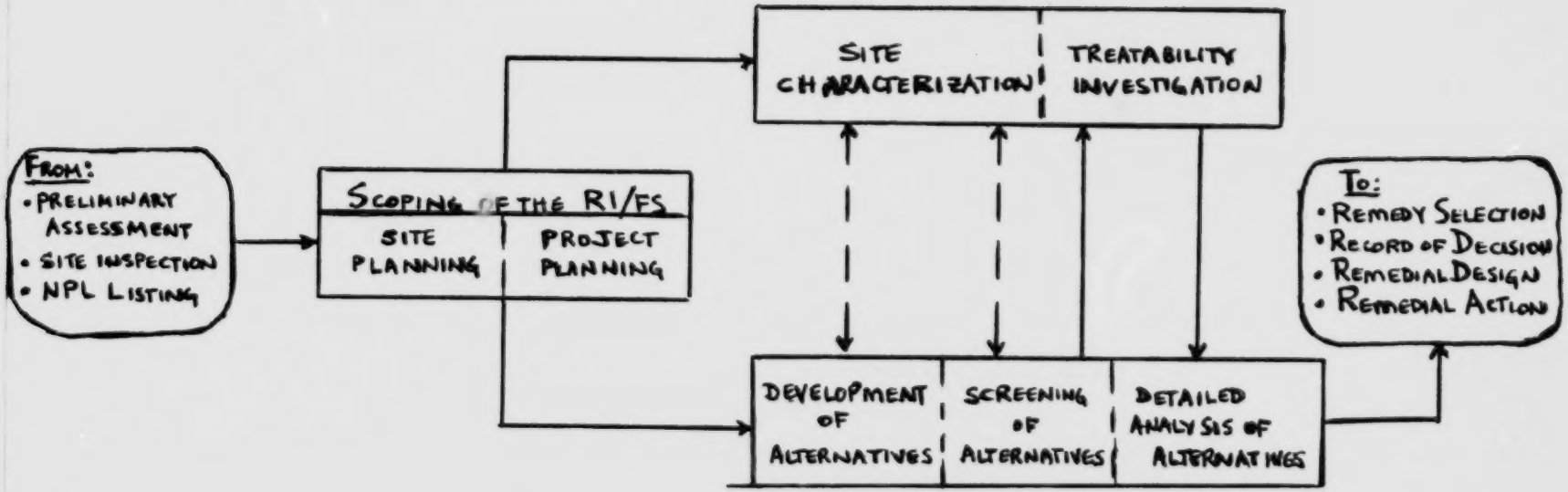
The development and selection of remedial actions and resultant cleanup levels are established through the Remedial Investigation (RI) and Feasibility Study (FS) process (also known as the RI/FS). The RI/FS consists of several major components: site characterization, risk characterization, development of remedial action objectives, and development and selection of the remedial action alternatives. A schematic of the RI/FS process and the interactions between its elements is shown in Figure B-1. The RI/FS process, being interactive, is one where the individual steps in the process are not discrete but consist of parallel and simultaneous operations which are interdependent, with ongoing reevaluation or adjustment in the light of any new information.

Site characterization is the initial phase where the nature of the contamination is assessed, including the identification of specific contaminants (chemicals), their concentration and the extent (volume or area) of contamination. Once information is available on the level of contamination, the risk characterization can begin. The risk characterization consists of the baseline public health evaluation which utilizes information from the site characterization activities.

Elements of the baseline risk evaluation are: the selection of indicator chemicals (when there is an extensive mixture of contaminants and chemical-specific ARARs do not exist for <u>all</u> chemicals present), which are those chemicals which represent the greatest risk; exposure evaluation (concentration at point of contact with exposed population or individual), taking into consideration all routes of exposure, not just that of the media of contamination; intake evaluation (actual amount taken into the body); toxicity assessment (taking into account chemical specific toxicological properties) and risk characterization, which is the final overall estimate of health impacts incorporating all the above variables.

Once the risk has been characterized target concentrations for the various chemicals in evironmental media are then

FIGURE B-1: SCHEMATIC REPRESENTATION OF RI/FS PROCESS



SOURCE: EPA, OERR, GUIDANCE FOR CONDUCTING REMEDIAL INVESTIGATION AND FEASIBILITY STUDIES UNDER CERCLA, DRAFT, MARCH 1988

developed, intended to reduce the overall risk to an acceptable level. These target concentrations are determined by apportioning the risk among chemicals present such that the overall risk falls within the acceptable range. The actual method of apportionment depends on the professional judgement of the individual(s) preparing the assessment and may or may not be rigorous. The apportionment may be as simple as an across the board reduction in all chemicals or may utilize greater detail where specific chemicals are targetted for greater reductions.

Acceptable risk can be met in one of two ways; (1) by meeting ARARs (assumed to be established on the basis of individual protectiveness) if they exist for all chemicals present, or (2) meeting a prescribed risk range. This risk range is designated as an overall risk of between 10^{-4} and 10^{-7} excess individual cancer risk per lifetime of exposure to the carcinogens present, and below the acceptable intake for chronic exposures to non-carcinogens. Assuming the existence of a threshold for non-carcinogens, acceptable intake for chronic exposures is defined as the highest human intake of a chemical that is not expected to cause adverse effects from long-term (lifetime) exposure.

In essence, the target concentrations define what is "clean". It should be noted that this target concentration is calculated for the point of exposure and neither specifies environmental concentration nor how this exposure reduction is to be met. The resulting environmental concentration may range from an extrapolated value based on the target exposure concentration (e.g., removal) to one where no change is made in the environmental concentration due to practical and technical concerns and exposure reduction (e.g., capping) is the focus of the action. The specific remedial action depends on many factors including exposure, environmental, technical, and economic considerations, which may have their own site specific differences and influence the selection of a remedial action. For this reason, variability in residual contaminant concentrations among different cleanup actions can be expected.

The next phases are the development, analysis, and selection of alternatives. Remedial action objectives take into consideration the contaminant, exposure route and receptor, and acceptable contaminant level or exposure level (i.e., target concentrations based on risk factors or ARARs). Response actions are then developed by medium based on the extent of contamination (volumes or areas). It should be emphasized that reducing exposures to acceptable levels may not always be achieved by actually cleaning up a specific medium to a given level but may focus instead on controlling the exposure by limiting migration.

Technologies are initially evaluated based on the technical feasibility for a particular medium and/or contaminant. Process options are selected to represent all technology types and then are assembled into alternatives to represent a range of treatment and containment combinations. Technology types include treatment, source control, and removal. In the early phase the technologies are evaluated with an emphasis on technical feasibility and effectiveness (ability to handle the type and quantity of contamination) and, to a lesser degree and only in a relative comparison, (administrative) implementability and cost.

Once alternatives have been developed and subjected to an initial screening, an in-depth screening of the alternatives can be carried out. The objective of the screening is to narrow the list of alternatives based on their effectiveness, implementability and costs. While the list of alternatives is narrowed, the range of options, (treatment, removal, and containment) is preserved. All alternatives are evaluated to insure that protectiveness will be achieved by reducing exposures.

Screening requires further definition of size and configuration of the actions, time frame in which the action will achieve it's remedial goals, rates of flow through the treatment process, spatial requirements, distances of transport to disposal, technologies, and required permits and imposed limitations. Further defined alternatives are screened on the basis of effectiveness (of protectiveness and reduction in toxicity, mobility, and volume), implementability (technical and administrative), and costs (estimates to an accuracy of +50%/-30%).(RI/FS Guidance Document, 1988) The result of this screening is to reduce the number of alternatives to a manageable number for detailed analysis.

Detailed analysis of the alternatives consists of further definition of each alternative, an assessment and a summary of each alternative against the nine evaluation criteria, see Table B-1, and a comparative analysis among the alternatives to assess relative performance with respect to each evaluation criterion. Because this is an interactive process, continuing redefinition of the alternatives may be required on the basis of new infomation (e.g., results of treatability study, new analytical results etc...).

The alternatives are analyzed on the basis of the nine evaluation criteria discussed in the previous section. The criteria and a brief description are presented in Table B-1. The result of this process is ideally the selection of the best option which balances the nine evaluation criteria and

Table B-1
Nine Evaluation Criteria

short-term effectiveness: primarily concerned with potential impacts during the construction and implementation of the remedial action, concerned with the protection of the community and workers, environmental impacts from implementing the alternative, and time until response objectives are achieved;

long-term effectiveness and permanence: potential impacts after implementation and remediation objectives have been met, magnitude of residual risk, adequacy of controls to manage residues and remaining untreated wastes, reliability of controls, and to assess the potential for future failures and the need for replacement;

reduction of toxicity, mobility or volume: performance of the specific technology: treatment process utilized, degree of expected reduction expressed as percntage of reduction, irreversibility of treatment, quantity of treatment residuals remaining;

implementability: both technically and administrative feasibility, details of construction and operation, reliability of technology, ease of adding additional remedial measures, monitoring, approvals and permits, availability of services, materials, equipment, adequate off-site treatment, personnel, and competitive bids;

cost: capital, operation and maintenance;

compliance with ARARs: how compliance is achieved or waiver required and justification for such a waiver;

overall protection of human health and the environment: meet the required risk range, endangered species and their critical habitiat, ecosystem; and

state acceptance and community acceptance: local concerns, issues, finances

other considerations. Some of the parameters are potentially competing, such as cost-effectiveness and the promotion of permanent treatment. Maximization of any one parameter would open the door for criticism on the basis of any other specific criteria. While each criteria has its

own value (or utility) it should be emphasized that this process intends to find a balance between these considerations and is not meant to maximize any particular parameter. In other words, it might be viewed as maximizing overall utility.

The nine evaluation criteria have been grouped as to their relative importance. OSWER staff have made informal groupings of the criteria into a hierarchy of four categories; threshold, primary, primary balancing, and modifying criteria. The threshold criteria include those elements required by law which must be met, protective of human health and the environment, and the attainment of ARARs (or a waiver if not feasible). The primary criteria are the next most important are those which are listed as specific considerations by statute and include reduction of toxicity, mobility and volume, implementability and cost. The primary balancing criteria include both long- and short-term effectiveness, and the final group, the modifying criteria, include state and community acceptance, are used as a final distinguishing characteristic. The grouping of the criteria gives an indication of relative importance. However, no absolute ranking or weights for the criteria are established. Therefore the balancing of criteria or maximization of overall utility must be done in an informal manner.

C. ROLE OF RISK ASSESSMENT IN EXISTING REGULATIONS

With the use of target concentrations and ARARs in the selection of remedial actions, risk assessment is being implicitly used as the basis of selection for determining final cleanup levels at the sites of CERCLA remedial actions. The effectiveness of this method and the possibility of improvements are contingent on the value of the risk assessment practices on which they are based. Therefore below I will first give a brief description of the conventional risk assessment practices presently in use and will critique them in a subsequent chapter.

1. Risk Assessment Guidelines

Much of the risk assessment which is carried out in the federal government is based on the National Academy of Sciences' Risk Assessment in the Federal Government: Managing the Process.(1985) The basic elements of risk assessment include hazard identification, dose-response evaluation, exposure assessment and risk characterization. Specific details of risk assessment methodologies used by the EPA are found in the Risk Assessment Guidelines of 1986. These guidelines are divided into five sections: Guidelines for Carcinogen Risk Assessment; Guidelines for Mutagenicity Risk Assessment; Guidelines for Health Risk Assessment of Chemical Mixtures; Guidelines for the Health Assessment of Suspect Developmental Toxicants; and Guidelines for Estimating Exposures. All risk assessment work for health impacts carried out within the EPA is required to use the above guidelines.

For simplicity, health risks can generally be divided into carcinogenic and non-carcinogenic effects. The EPA assumes in the absence of contrary data that carcinogenicity is a toxic response that does not have a threshold and therefore is also referred to as a non-threshold effect. A threshold is defined as a level below which no effects are observed. The assumption of no threshold is an example of a conservative assumption, where in the absence of sufficient contrary data the most reasonably conservative alternative is chosen. This conservative approach dominates the EPA risk assessment and regulatory philosophy on the basis that it inherently provides an additional degree of confidence that protection will be assured. Conversely, non-cancer endpoints are assumed to have a threshold characteristic.

In general, because of the non-threshold assumption carcinogenicity is usually the driving force in a regulatory decision for substances having evidence of carcinogenicity. Mutagenicity and developmental toxicants generally play a less significant role in such decisions but are considered

if available information exists. For simplicity, only the basics for carcinogenic and non-carcinogenic assessments will be presented here.

1.A Non-threshold Contaminants:

Guidelines concerning the assessment of carcinogenicity are outlined in the Guidelines for Carcinogen Risk Assessment (51 FR 33992), one of the five guidelines cited above in the Risk Assessment Guidelines, and summarized below. The initial phase of the risk assessment process is hazard identification, which includes a review of available information on: physical-chemical properties and routes and patterns of exposure; structure-activity relationships; metabolic and pharmacokinetic properties; toxicologic effects (short-term tests); long-term animal studies (at or near the maximum tolerated dose level, MTD, which is essentially a Lowest Observed Adverse Effect Level, LOAEL for a subchronic study); and human studies. From available information in animal studies a weight of evidence scheme is developed. For any given animal study the strength of evidence that an agent is potentially carcinogenic for humans increases:

(1) with the increase in the number of tissue sites affected by the agent;

(2) with the increase in the number of animal species, strains, sexes and number of experiments and doses showing a carcinogenic response;

(3) with the occurrence of clear-cut dose-response relationships as well as a high level of statistical significance of the increased tumor incidence in treated compared to control groups;

(4) when there is dose-related shortening of the time-to-tumor occurrence or time to death with tumor; and

(5) when there is a dose-related increase in the proportion of tumors that are malignant.

Once all information is collected on individual studies and evaluated an overall weight of evidence classification is made based on available data. Evidence of possible carcinogenicity in humans comes primarily from two sources: long-term animal studies and epidemiologic investigations. There are three major steps to characterizing the weight of evidence for carcinogenicity in humans:

(1) characterization of the evidence from human studies and from animal studies individually,

(2) combination of the characterizations of these two types of data into an indication of overall weight of evidence, for human carcinogenicity, and

(3) evaluation of all supporting information to determine if the overall weight of evidence should be modified.

The EPA classification scheme is, in general, an adaptation of the International Agency for Research on Cancer (IARC, 1982) approach for classifying the weight of evidence for human data and animal data. A comparison of the EPA and IARC classification schemes is presented in Table C-1.

Table C-1
Weight of Evidence Categorization in Carcinogenic Risk Assessment

Group I: Known human carcinogens (sufficient epidemiological evidence)	Group A: Human (or animal) carcinogen (sufficient epidemiological (or experimental) evidence to support a causal association)
Group IIA: possible human carcinogens, limited epidemiological evidence	Group B: Probable human carcinogen B1 - limited epidemiological evidence
Group IIB: possible human carcinogens, limited epidemiological evidence	B2 - inadequate or no epidemiological evidence; sufficient experimental evidence
Group III: inadequate evidence for human carcinogenicity	Group C: possible human carcinogen limited or inadequate experimental evidence
Group IV: no evidence	Group D: not classified due to inadequate evidence
	Group E: evidence of non-carcingenicity based on adequate experimental data or adequate experimental and human data

Source: Silbergeld, Ellen K., Hazardous Waste & Hazardous Materials, Vol. 4, No. 2, 1987, p. 145

Agents which are judged to be either Group A or Group B are regarded as suitable for quantitative risk assessments. Agents classified as Group C will generally be regarded as suitable for quantitative risk assessment, but judgements in this regard may be made on a case by case basis. Agents that are judged to be in Groups D and E would not have

quantitative risk assessments performed.(Guidelines for Carcinogen Risk Assessment, 1986) It should be noted that those compounds which are designated as E would therefore only be of concern as to their non-carcinogenic effects.

Dose-response assessments utilize available data from animal and human studies, selected by criteria outlined in the guidelines, and applicable mathematical extrapolation models from high to low dose. Once the data have been selected, a cancer potency value is derived from the linearized multi-stage model with upper 95% confidence limits on risk. Many other models exist such as the one-hit, multi-hit, Weibull, logit, or probit models, and extrapolation from high to low dose utilizing these models may result in differences in the estimate of risk at low doses by several orders of magnitude. While there is insufficient knowledge of the mechanisms of carcinogenicity to imply that any model is superior to the others from a regulatory perspective, the linearized multi-stage model is the EPA's convention, unless there is specific evidence to the contrary, because it is believed to yield consistency in how these estimates are derived and represent a plausible upper limit estimate of risk.

The exposure assessment either utilizes a series of accepted values (e.g., water consumption, body weight, etc...) or, if available information exists, carries out an analysis for a particular exposure scenario. Unless there is evidence to the contrary, the cumulative dose received over a lifetime, expressed as an average daily exposure prorated over a lifetime, is recommended as an appropriate measure of exposure to a carcinogen. Once information is developed from the exposure assessment (accounting for concentration and population exposed), expressed in terms of mg/kg/day intake, the cancer potency value can be applied to determine the upper bound estimate of the number of excess cancers which would be expected per ten thousand (10^4), hundred thousand (10^5)or million (10^6) persons in the exposed population. The actual risk is assumed to not likely exceed this upper bound estimate and may, in fact, be lower.

1.B. Threshold Contaminants:

Unlike non-threshold contaminants where effects are quantified for any concentration of exposure (intake), non-carcinogenic effects are approached on a threshold basis and set to insure against adverse efects. A threshold (concentration and implied exposure or dose) is quantified. Any daily exposure to the human population (including sensitive subpopulations) to this level would not likely result in any appreciable deleterious health effects during a lifetime. These threshold values are calculated for non-cancer toxicity endpoints. The actual terminology for this threshold value differs among programs and covers both

chronic and acute exposures. Various terms include; Reference Dose (RfD), Acceptable Daily Intake (ADI), Acceptable Intake for Chronic Exposures (AIC) and Acceptable Intake for Subchronic Exposures (AIS).

Threshold values are calculated by first identifying no-observed-adverse-effects levels (NOAEL), no-observed-effects levels (NOEL), lowest-observed-adverse-effects levels (LOAEL), or lowest observed effects level (LOEL), all in units of mg/kg/day, from human and animal studies of appropriate exposure duration and concentration. Calculations are then made using uniform exposure assumptions (e.g., consumption of 2 liters of water consumed per day, 70 kg body weight for an adult and 10 kg for a child) to calculate concentrations in the media which would correspond to the doses identified (i.e., NOAEL, NOEL, etc...).

Frequently there is limited data for human exposure to a particular substance of interest, either of adequate duration or concentration. Therefore extrapolations among species, routes of administration and dose must be made to account for such data gaps. Compensation for these extrapolations occurs through the use of uncertainty factors. Uncertainty factors are applied to the best estimate using existing information, dividing the best estimate by the appropriate uncertainty factor. This uncertainty factor is made up of several components and can vary from 10 to upwards of 10,000. A list of uncertainty factors is shown is Table C-2. A factor of 10 is used for each of the uncertainty factors that apply and multiplied

- -

Table C-2
Uncertainty Factors
(10X used for each that apply)

NOAEL is used from a human study (believed to compensate for intraspecies variations of response.

NOAEL is identified in an animal study (to compensate for interspecies variation),

Dose extrapolations are used (subchronic exposure used to estimate a chronic exposure),

LOAEL in a human study is used instead of a NOAEL.

NOTE: An additional uncertainty factor of 10 or less is used to account for differences in absorption, study quality deficiencies or extrapolation from different routes of exposure.

- -

together to get an overall uncertainty factor. The NOAEL, LOEL or related value are then divided by the appropriate uncertainty factors to obtain the the acceptable concentrations, (e.g., ADI, RfD).

2. Program Specific Use of Risk Assessment

While there may be subtle differences (primarily in terminology) among offices or programs in EPA, the basic risk assessment methodologies do not differ. While the assessment of risk is consistent between programs, how the information is used within the programs (e.g., to establish standards) is a major source of difference and is dependent on the focus or mandates of each individual program. For many programs the risk assessment is only a portion of the overall information needed. Many programs will also include cost and/or technical feasibility as additional considerations in decision making.

To date, all EPA program standards do not consider synergism or additivity. Standards are established as if the compound and route of interest was the only exposure of concern. However, if there are sufficient data to suggest that specific compounds of concomittant exposures would have synergistic effects then the acceptable standard would be lowered to account for that. Unless contrary evidence exists, standards are developed under the assumption that the regulated chemical and the route of interest are the only contaminant to, and route by which an individual is exposed.

As Section 121 of SARA specifies, the selected remedy must attain or exceed levels established in applicable or relevant and appropriate requirements (ARARs). The most common ARARs associated with Superfund actions are those directed at water, either for consumption or in the aquatic environment and include Maximum Contaminant Levels (MCLs), Maximum Contaminant Level Goals (MCLGs) and Water Quality Criteria (WQC). To fully understand the role of risk assessment as it relates to establishing clean levels one must understand the basis for the ARARs. Therefore a brief overview of the health and risk basis for the most common potential ARARs follows.

2.A. Drinking Water Standards (MCLGs and MCLs)

MCLs and MCLGs are standards for drinking water and are intended to be protective of the the human population under conditions of normal consumption. The development of enforceable standards for drinking water consists of two parts: setting the maximum contaminant level goal (MCLG), a health-based goal that is non-enforceable, and the maximum contaminant level (MCL), which is the enforceable standard under the Safe Drinking Water Act (SDWA). For non-cancer

effects the MCLG are derived from the Reference Dose, RfD (formerly called the Acceptable Daily Intake, ADI), which as explained above represents the level of lifetime exposure at which no adverse health effects would be expected.

The RfD is adjusted for a 70-kg adult consuming 2 liters of water a day assuming 100% of exposure to a substance is from drinking water resulting in the Drinking Water Equivalent Level (DWEL). However, as drinking water is not the only source of exposure to contaminants an additional apportionment factor is used to reflect the actual exposure drinking water represents for a compound. When actual exposure data are not available it is assumed that exposure to drinking water represents 20% of the total individual intake and the DWEL is adjusted accordingly. The resulting number is considered to be the lifetime Health Advisory (HA), is equivalent to the MCLG, and represents the health goal in the regulatory development process under SDWA.

Health advisories for non-cancer effects also are developed for less-than-lifetime exposures, and are meant to be used as guidance in emergency situations (such as dealing with emergency spills or contamination). Less-than-lifetime HAs are calculated in much the same manner as the lifetime RfD by using available data and applying uncertainty factors. However, less-than-lifetime HAs are usually derived from studies involving short-term, continuous exposures followed by an observation period. Additionally they differ from MCLGs (lifetime HAs) in that 100% of an individual's intake is assumed to come from drinking water. These less-than lifetime HAs may be considered acceptable for periods of up to a few years.

As carcinogenesis is assumed by the EPA to have no threshold, the MCLG is established at zero for all known or probable carcinogens. The MCL represents a level of acceptable or reasonable risk for drinking water contaminants and is set as close to the MCLG as feasible. Feasibility is determined in part by costs of achieving end of the pipe concentrations for commercial (>25 connections) water systems, and considers both treatment technologies and practicable detection limits.

Once the MCLG is developed a review of field tested technologies is made. This review encompasses economics and performance achievability. Information is entered into a cost model and technologies are chosen which most effectively achieve levels at or as near as possible to the MCLG for a reasonable cost. Performance standards are developed with full peer review and have been carried out on a compound-specific basis in the past.

Another factor which influences the MCL is the practicable quantification limit (PQL), the practicable level of

detection for an average water treatment plant laboratory considering costs, technologies and available skills. This does not represent the state of the art in analytical chemistry, but what is considered commonly practicable. Standards are set at the PQL (if the PQL exceeds the MCLG) or above the PQL (when the MCLG exceeds the PQL). The SDWA mandates that a review of MCLs be carried out every three years and adjustments made for the emergence of new technologies on a contaminant-specific basis. In theory as new technologies are developed the MCL will approach the MCLG. Variances are possible if designated Best Available Technology (BAT) does not achieve the levels established as the standard.

2.B. Water Quality Criteria (WQC)

Water Quality Criteria are non-enforceable guidances issued under the jurisdiction of the Clean Water Act to aid states in developing their own specific laws to protect human health and aquatic life. They consist primarily of scientific documentation. The states can factor in economics and technology and other site specific consideratons (such as indigenous species and existing uses) in establishing their own standards.

WQC for human health effects are calculated much the same as MCLGs. In fact, WQC are essentially equivalent to MCLGs. One major difference is that to date the WQC does not apportion the risk to account for other sources (as is done in going from DWEL to HA, MCLG). Existing data are reviewed, a NOAEL or LOAEL is identified for non-cancer effects and the appropriate uncertainty factor is applied. For carcinogenic effects a range of 10^{-5} to 10^{-7} lifetime probability of cancer death from exposure is presented as a guide to what is appropriate, though the actual recommendation is zero, much as in the MCLG.

WQCs differ from MCLGs and MCLs with regard to human health in that the WQC considers bioaccumulation of a substance in the tissues of aquatic life and intake from the consumption of that fish in the exposure scenario, which is not considered under drinking water standards. Minimum data requirement for the WQC is at least one valid long-term feeding study in mammals. Regulations under the CWA specifically allow estimates of allowable intakes for water consumption to be derived from data concerning health effects induced by inhalation. However, while this is allowed it has never been used as the sole source for a standard.

WQC for aquatic life is a standard dealing with protection of a non-human effect and/or the ambient environment. Studies are made to determine safe levels for aquatic life. A range of species type must be included in the studies;

including a sensitive species such as trout, and include both vertebrates and invertebrates. The biological endpoints of mortality, growth reduction and reproductive function are examined. Protection is established as the standard which would insure no effects for 95% of the species tested. Field data if available are also incorporated into the standard. In general, the data used in setting standards associated with aquatic life are felt to be more representative and accurate than the data associated with the prediction of human health effects as they are obtained in a more controlled setting.(Category 1) This is supported by the fact that there is less use of extrapolation and fewer uncertainty factors in the development of an aquatic life WQC than those used in establishing human health standards.

3. Environmental/Ecological Risk Assessment

The capacity for environmental and ecological risk assessment within the EPA is still at a stage of early development when compared to that for human health assessments. At present, environmental effects are handled qualitatively. Field data and observations are typically used as the basis of decisions as opposed to actual modelling. While models do exist the guidance for their usage is limited and it is left to professional judgement whether such models should be used.(RI/FS Guidance, 1988)

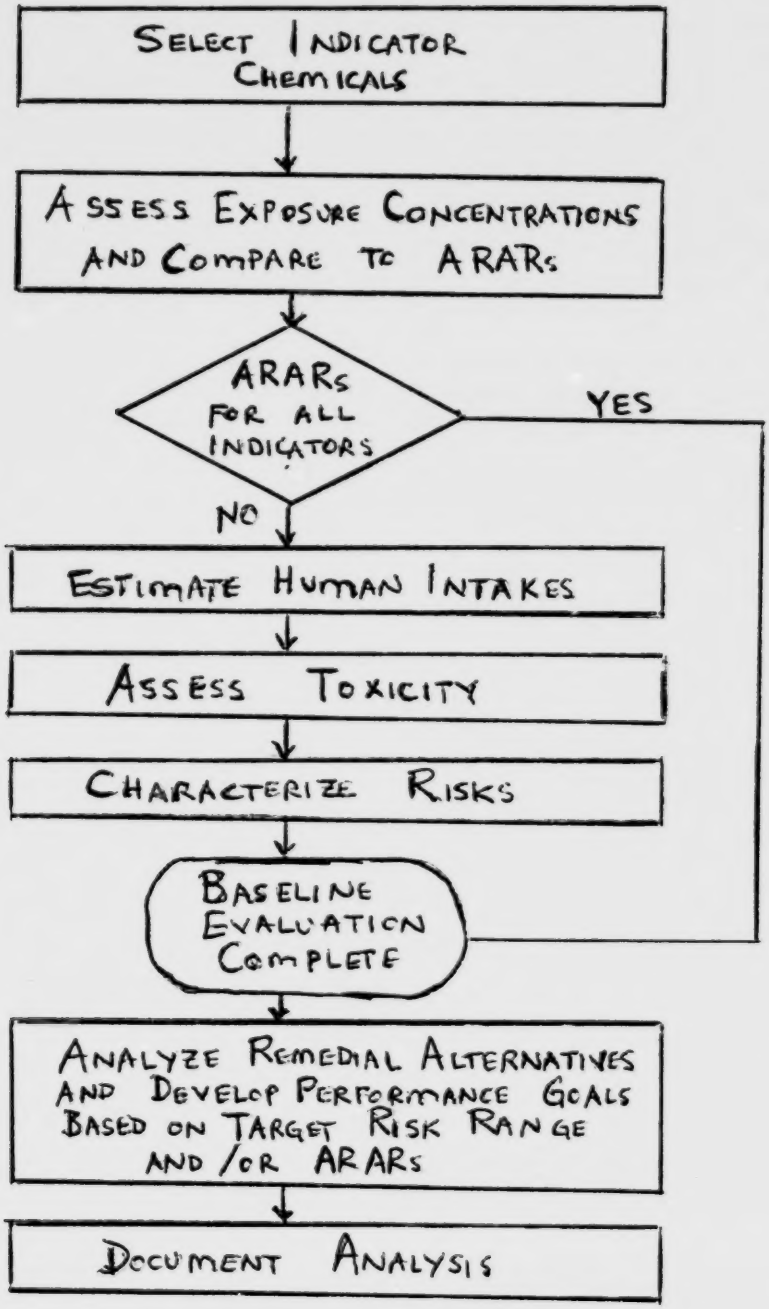
The EPA has identified this as an area of significant concern and has undertaken a major effort to develop guidance on ecological effects evaluation. This guidance is expected to be completed and distributed sometime in 1989.

4. Superfund Public Health Evaluation

Risk assessment in the Superfund program is dictated by the procedures outlined in the Superfund Public Health Evaluation Manual (SPHEM). The risk assessment or public health evaluation process as discussed in the previous chapter is outlined in Fig. C-1. Once information is received from the site characterization activities, indicator chemicals (when there is an extensive mixture of contaminants and chemical-specific ARARs do not exist for all chemicals present) are selected. Using environmental transport models exposure concentrations are calculated. Intake and dose are assessed and along with toxicity assessment the risks are characterized. Two existing databases, Public Health Risk Evaluation Database (PHRED) and Integrated Risk Information System (IRIS), are the primary sources for toxicity information. In addition and independently the Agency for Toxic Substance and Disease Registry (ATSDR) conducts a health assessment. While the information used by those in the EPA includes animal

FIGURE C-1

FLOWCHART OF THE SUPERFUND PUBLIC HEALTH EVALUATION PROCESS



SOURCE: EPA, OERR, SUPERFUND PUBLIC HEALTH EVALUATION MANUAL, DECEMBER 1986, EPA/540/1-86/060

Reproduced from best available copy.

studies, ATSDR prefers to use primarily human data and related methods such as structure activity relationships and pharmacokinetic modeling.

Once the risk has been characterized target concentrations for the various chemicals in evironmental media are then developed, intended to reduce the overall risk to an acceptable level. These target concentrations are determined by apportioning the risk among chemicals present such that the overall risk falls within the acceptable range, or by attaining ARARs (for which specific risk methodologies were previously discussed). In essence, the target concentrations define what is "clean". These target concentrations are then used as goals to design remedial actions.

There are no unique risk assessment techniques in the Superfund program when compared to other programs within the EPA, as discussed above. Rather existing standards, databases, models and toxicity information are used, and SPHEM is the procedure which organizes these techniques and protocols into a uniform formal framework.

D. CRITIQUE OF EXISTING METHOD OF REMEDY SELECTION AND ESTABLISHING CLEANUP LEVELS

Completeness of Conception (Breadth of Analysis)

The RI/FS process, as briefly outlined in Section B, provides an excellent guidance for complete problem definition and identifying concerns. The RI/FS process serves to bring together information from unrelated disciplines and provide a means of allowing direct comparison. It provides a framework to guide site characterization, to determine the nature of the hazards, the type and extent of contamination and physical characteristics of the site, the potential for mobility, transport and resultant exposure. Specific factors addressed include geology, hydrology, geography, ecology, and population. Throughout my research it was found to be uniformly agreed upon that the RI/FS process as designed is adequate in scope. (Categories 1-4)(EDF, 1988, OTA, 1988)

A good indicator as to the completeness of conception is the ability to incorporate the extremes of subjective judgement and empirical or "pure" evidence. The RI/FS process as defined is not limited in its use of available information or data. The process is flexible enough to accomodate both subjective judgement (as in the choice of fate transport models) and data or evidence (such as those from site sampling). This flexibility to allow a broad range of inputs is one of the strengths of the RI/FS process.

While the RI/FS process is flexible, allows an expanded problem definition, there are still elements which are lacking. The definition of risk is not complete, relying on single upper bound point estimates, and will be discussed in greater detail in the next chapter. The process of remedial action design and selection is dominated by human concerns or perspectives. The ethical perspective or the rights of the environment to protection are inadequately addressed.

While the list of potential individual factors influencing CERCLA actions is fairly complete in the RI/FS, the level of analysis each parameter is afforded after identification is left to the professional judgement of those involved in the analysis. The analysis for any individual site is theoretically dictated by information obtained regarding individual site characteristics using the framework outlined in the RI/FS as a guide or checklist.

Rigor Of Thought (Depth of Analysis)

As stated above the RI/FS process allows professional judgement to dictate the level of analysis of any particular parameter. The flexibility of the process allows one to structure the analysis or decision based on available

information or data. The available evidence is designed to dictate the basis of any decision, not just a prescribed decision formula. This intent has been supported in the citation of several commendable examples of the adequate use of the RI/FS process (OTA, 1988, EDF, 1988).

The RI/FS guidance provides accepted values or ranges of values for parameters to be used in analysis. The level of sophistication or coverage of any particular parameter (e.g., geology, transport etc...) is limited in depth. Therefore the depth of analysis is dependant on both the skills and judgement of the technical staff preparing the analysis. Frequently a range of accepted values are assumed as opposed to the use of potentially more relevant values.

One major reason why the depth of information for any parameter is limited is the limited information to date. The Superfund program can still be considered to be in its infancy and there is little information or experience to draw from. It should be noted that the EPA recognizes this concern and is making attempts to greatly improve the guidance and to develop a better overall information base, incorporating and disseminating information and experience as it becomes available.

While the inclusion of ARARs represents a means of incorporation of related experience, the appropriateness or relevance of the experience and/or standard is not adequately considered. While it may be useful to include these standards one must consider that all previous EPA programs (e.g., water and air) have been media specific with specific mandates or circumstances. The incorporation of ARARs may be integrating regulations with potentially conflicting mandates. The standards of the other programs (water and air) have been developed in response to individual legislative mandates and in relative vacuums from considering other regulations or standards. These individual mandates vary from the protection of the most sensitive individual with an adequate margin of safety regardless of cost (as in the air program) to regulations which are specifically required to factor cost into the equation (as in the water program).

In addition to the mandates of the regulation, there are differences in the goals, mechanics and affected parties for such regulations. For example, the air program concerns itself with ambient concentrations but it's standards and control strategies are frequently targetted at specific industries or processes. In contrast MCLs under the SDWA are concerned with end of the pipe concentrations for commercial water systems with greater than 25 connections and control strategies are tied directly to the water treatment and distribution system.

Regulations also differ in the mechanism of the regulation. Some standards developed only have concentration-based standards based on health, such as NESHAPs, MCLGs, and WQC. Others have a technology-based component, such as MCLs and the land ban requirements of RCRA. Typically standards have a component of technical feasibility incorporated into them. However, these technical considerations while applicable to say a drinking water treatment and distribution system, may not have the same relevance to a typical situation for CERCLA, like groundwater treatment. Because of these different objects of regulations and feasibility tests this mix of different goals and foundations of regulations can lead to inefficiencies.

By the inclusion of the MCLGs as a goal, Superfund would be required to clean up an unused aquifer to a non-enforceable MCLG level while the water program is not even required to meet these standards for existing water consumption. This has created an inconsistency in the eyes of many involved in the program.(Category 1 and 2)

It should be noted that the RI/FS does address this by the availability of waivers from attaining specific ARARs. A chemical-specific ARAR may be waived if it can be shown that (1) there would be greater risk to health or envirnoment by using an action to meet the ARAR, (2) technical impracticability, or (3) fund balancing (excessive costs required to meet the ARAR). The determination is made for each ARAR with CERCLA actions. However, it is the opinion of the author that in addition to the typical considerations of ARAR determinations on a site specific basis with respect to relevant and appropriateness, overall CERCLA-wide determinations should be made as to whether the actual standard (ARAR) is applicable to CERCLA situations.

Standards are chosen balancing available health, economic and technical data. Each of these elements has a degree of uncertainty. A given level for a standard may be acceptable when balancing all information for a given application (e.g., end of the pipe concentrations for drinking water). However, when applied to a different situation the uncertainties associated with this different application may lead one to make a different selection for an acceptable standard.

The RI/FS process like most technical analyses generates numerical estimates. These numerical estimates are not representative of nature but are only estimates. The value or accuracy of these estimates is dependant on how they were derived, reliability of available data and/or the model used. Numerical estimates have associated uncertainties. However, frequently numbers are presented or used in such a manner as to give them an appearence of being an absolute. This is also the case in Superfund, where site

characteristics or risk estimates are given as single values. The use of these estimates in such a manner is representative of a lack of depth and can lead to inefficiencies or misguided actions. This will be discussed in greater detail in the next chapter.

For example, an estimate of risk of 10^{-5} lifetime probability is not actually a single value but a range, which based on conservative assumptions is not likely to be exceeded and may be in fact lower. However, the public, or even policy analysts, may interpret that estimate as the actual value (i.e., the risk *is* 10^{-5}). This perception of an absolute can lead to a misinterpretation or perception and invalid expectations.

Risk assessment and it's results are particular sensitive to such misinterpretation. Not surprisingly a criticism has been voiced that risk estimates have been improperly communicated to the public, hindering credibility and resulting in unattainable and unrealistic expectations.(Category 1, 2, and 4) Too much emphasis has been placed on the numbers, losing sight of their meaning, or relative importance, or origin. This too will be discussed in greater detail in the next chapter.

Consistency of Application

Under the RI/FS process the baseline health risk determinations follow a fairly consistent format outlined in the Superfund Public Health Evaluation Manual. The existing methodology is very clear in the definition of existing site conditions and the concommitant health risks. Utilizing the information of the baseline (existing) conditions, derived from uniform method of analysis, allows a comparison and prioritization among sites.

The definition of protectiveness is the most consistent element of the Superfund program. Protective measures are defined as those attaining or exceeding ARARs for all compounds, if available, or within a range of 10^{-7} to 10^{-4} excess individual cancer risk per lifetime of exposure to carcinogens and below the acceptable intake for chronic exposure for non-carcinogens.

As stated previously professional judgement plays a major role in the direction and depth of analysis. While the flexibility afforded because of this may be neccessary in dealing with site-specific differences it does have the potential to result in inconsistencies. In fact, inconsistency may be the one most common complaints of the Superfund experience.(EDF, 1988, OTA, 1988) The "every site is unique" argument to support this flexibility is believed to have been abused.(EDF, 1988, OTA, 1988)(Categories 1-4) Some have felt that this philosophy is the source of many

delays in the initiation and completion of remedial actions.(Category 2 and 3) However, sources directly involved in the process feel that engineering data collection and design are actually the time limiting step.(Category 1)

The exposure assessment is considered to be the weakest link in the process and believed to be inconsistent and the target of much manipulation.(Category 1 and 3) The SPHEM provides a range of acceptable values to be used for exposure parameters but lacks sufficient guidance in their validity or usage. As a result there is a variability of what values are used; typically consultants, states, and regions all use values which they find acceptable with little or no communication between these parties as to their acceptability.

It should be noted that this opinion is not limited to EPA critics alone, but is also present in the agency itself. As a result the agency has embarked on a program of improving the quality and consistency of such analysis. This will occur through the revisions to the SPHEM and SEAM, the development of a review mechanism for risk assessments, better communication between regions, and improved training and guidances.

The problems of inconsistency are not limited to the risk assessment portion of the process. Other areas such as determinations of technical and economic feasiblity have varied. For example, cost assumptions used for calculating overhead have varied between regions which has influenced the selection of remedial technologies.(EDF et al, 1988, OTA, 1988) Criticisms have also been made that the presentation of assumptions, data, and logic train have not been made in the RODs, thereby making a critical review of decisions difficult, and reducing public confidence in the program. For even if the decisions were made on a sound basis, the lack of presentation of the logic train in the RODs does not allow public scrutiny.(EDF et al, 1988)

The EPA has recognized this and has targetted making improvements in consistency between states, regions and consultants. Given the nature of the complex problem addressed by Superfund, flexibility is not essential, but the optimal mix of flexibility and guidance has not yet been reached.

Demonstrability of Goals

The problems associated with a Superfund cleanup are complex. Indicative of this complexity are the range of goals and the nine evaluation criteria developed to select remedial actions. These criteria represent a broad range of goals and as stated previously an action is selected which

is felt to provide the greatest balance between them. As there is no formal method of weighing these criteria (goals) there presently is no mechanism to determine if goals have actually been met. The only goal which is directly measurable is that of protectiveness, (with a hazard index of <1, carcinogen risk of 10^{-7} to 10^{-4}, and/or attainment of ARARs). The other goals (criteria) would have to be judged subjectively if they have been met and to what degree.

E. CRITIQUE OF EXISTING RISK ASSESSMENT

Completeness of Conception

The process of risk assessment within the EPA has developed into a principally mechanistic process following a prescribed set of rules. These rules are set out in the previously described Risk Assessment Guidelines. The EPA appears to have traded some of the more sophisticated elements of risk analysis for a more simple workable model. A major criticism of the EPA methodology is that simplifying the issues involved and not including all relevant concerns does not yield results which are indicative of the actual degrees of risk.(Category 1, 4)

Alternative models, thresholds for carcinogens, use of all available data, level of confidence determinations, consideration of other exposures, and population risk characteristics are not generally included in conventional risk assessments. The use of a maximum exposed individual as the determinant of risk, the use of the linear multi-stage model, and representation of risk as a single point estimate are examples of adopted simplifications. The use of these assumptions and simplifications are meant to represent a plausible upper-bound estimate of risk, which is a narrow depiction of risk.

Risk is not a single point estimate but a distribution or range of estimates of predicted outcomes based on applicable theories, models, and data. Existing methodology uses a single model with a common set of assumptions and lacks the ability to properly communicate the broader concept of risk. Risk assessments would be improved by the inclusion of the range of weighted estimates and an overall cumulative probability presentation of risk.

Uncertainty

All inferences of risk contain an element of uncertainty, in both the parameters chosen and in models employed for endpoint prediction. This is not limited to the health effects prediction in risk assessment. While it has been shown for some air studies that the health effects (dose-response) predictions have greater uncertainty than the environmental models (Amaral, 1987), the uncertainties of all aspects of the risk assessment process must be considered. The risk assessment process is a conglomeration of many steps consisting of many models and predictions, each with associated uncertainties. These models appear at many levels within the process, as indirect as those used in establishing standards (ARARs) to those directly associated with Superfund cleanup (exposure models, environmental transport models, etc...).

Site characterization is also subject to uncertainty. Sampling may not be representative if a hot spot or limited contaminated area was sampled and believed to be indicative of the site as a whole. Sampling is an expensive activity and therefore decisions are often made with limited information. The accuracy of the sampling information is very important, a real source of uncertainty and not adequately considered at present. No matter how extensive a model may be it is only as good as the data (sampling) which are placed into it. Therefore better representation should be made as to the quality of the data used in the overall risk assessment.

The simplified EPA methodology does not present uncertainties in any discernible manner because of its presentation of a single estimate using a single model. The aggregation of conservative values in risk estimation is believed to yield an overall "reasonable confidence" in the final estimate. This policy of risk estimates representing the upper bound with the actual value likely to be lower makes a final representation of uncertainty unnecessary. However, opinions have been expressed that more sophisticated techniques and quantified uncertainty analysis should now be required.(Categories 1-4)

Methods are available to quantify or better present uncertainty and should be integrated into the EPA methodology so as to allow determinations in overall confidence in any decision or estimate. At present one must rely on the policy of "reasonable confidence" in the risk estimate rather than determining the actual confidence in any decision.

Population Risk v. Individual Risk

Risk can be presented as either individual or population risks. Individual risks are those risks for any individual, usually calculated for the maximally exposed individual and presented in the familiar fraction (e.g., 10^{-6}). Population risks are the overall anticipated effects given distribution of susceptibility or response to contaminants within a population. These differences in susceptibility may be due to individual characteristics such as sex, age, predisposed conditions, and lifestyle, or demographic concerns such as distribution of population and industries.

Population risk is frequently incorrectly presented as the product of the individual risk and the total population. For that to be true susceptibility would have to be uniform. However this susceptibility is (typically log-normally) distributed which makes population risk much more complex, a fact which is frequently overlooked or not adequately

addressed.(Silbergeld, 1987) While each presentation of risk is useful by itself the two estimates together provide a superior description of risk.(Rodricks et al, 1987)

The EPA risk assessment methodology provides for the identification and consideration of sensitive individuals but lacks the orientation of determining overall predicted effects based on the characteristics of the exposed population. Population risks and overall cases should be emphasized more in the future.

Additivity and Synergism

The single-medium, single-compound orientation of the standards does not consider other routes or contaminants. Additivity and synergism while mentioned in the guidelines are not covered in any great detail.

Extrapolation of Existing Health Data

The EPA has made an effort to expand the relevant sphere of risk data by incorporating data and information from varied and applicable sources. These include the use of data from other sources such as animal studies, human (typically occupational) exposure studies, structure-activity relationships, and pharmacokinetic modeling.

While there is concern over the applicability of animal data to humans, there are several reasons to indicate sufficient similarities. These are: a uniform genetic code for all mammals, similar cell structure and chemistry, similar body temperature, and the fact that lung volume, respiratory rate, heart rate, blood flow rate, size of blood vessels and oxygen utilization are all functions of body weight which may be scaled between species. In addition to interspecies extrapolation, dose extrapolations are also done typically using the linear multi-stage model to allow the use of data and information from studies conducted at higher exposures than would typically be found at Superfund sites. The EPA has adopted a policy of the use of uncertainty factors to account for the problems with the use of these data.

Environmental or Ecological Risks

The intent of the Superfund remediation program is to develop remedies which are protective of the public health *and* the environment. While the methodology in the SPHEM is fairly extensive, it concerns itself only with human health effects. There is a general consensus that environmental and ecological hazards are not adequately addressed. (Categories 1-4) Risks exist for long-term environmental effects as well as human health and the existing risk assessment methodologies for environmental effects is presently limited.

Present eco-risk methodology is primarily a subjective analysis or determination of environmental effects and protection. The only real standard which is protective of non-human life is the WQC(aquatic) standard, which is protective of 95% of the species tested for both lethal and non-lethal (developmental) effects. There is a general lack of agreement on environmental data and models used within the agency, and therefore a database similar to that on health effects does not exist. The major point of contention appears to be the lack of concensus on what constitutues damage (i.e., endpoints to be estimated). While the EPA is taking strides to develop a comprehensive "eco-risk" program it is plausible to say that it is presently deficient.

Risk Perception

Both the use of extreme upper bound estimates and the presentation of risk as a single point estimate have severe effects on how those risks are perceived. While individuals implicitly accept risks and carry out subjective assessments in doing so, the perception of the risks is not always consistent with the reality of the estimates or experience.

Major infuences in perceptions of risk is whether it is of high consequence and voluntary or involuntary. High consequence events, regardless of rarity, are feared over more frequent but less catastrophic events. Involuntary risks are generally perceived as greater or more harmful than voluntary risks, probably because of the feeling of no control over ones's fate.

Perceptions of the risks associated with uncontrolled releases from chemical wastes suffers from both of the above, high-consequence and involuntary exposures. As a result, the fears associated with this issue influence both the public acceptability of EPA actions or inactions. As the risks associated with Superfund are susceptible to overexaggeration, the use of extreme upper bound estimates may be affecting the expectations of the public in an unwarranted manner.

Rigor of thought (Depth of Analysis)

As the EPA has taken a simplified approach to risk assessement to improve consistency, depth of analysis, like completeness of conception, had to be sacrificed.

Relevance of Uncertainty Factors

To integrate information from other sources such as animal studies uncertainty factors are used. These uncertainty factors are used without regard to the actual uncertainties

associated with the particular extrapolation of evidence. This liberal use of uncertainty factors can have deleterious effects on the quality of the final risk estimates. No inferrences can be made as to the accuracy of the final risk estimate or its overall uncertainty. The factor of ten used for each of these uncertainty factors is considered conservative but no studies have been made as to whether this overall approach is protective, unprotective or overprotective. A more sophisticated approach of actually looking at the variability of the data or nature of the evidence to determine the uncertainty in extrapolation and the overall risk estimates would be better.

Relationship between Uncertainty and Confidence

The use of uncertainty factors, while being protective, may actually skew the risk assessment and risk perceptions. Agents with good human data or extensive animal data and considered to be known human carcinogens do not have large uncertainty factors incorporated into them. Agents with little data have uncertainty factors sometimes as large as 10,000 incorporated into their standards. This could skew the concern from known carcinogens and focus attention, perception, and effort to those with little evidence of human effects. This may lead to lower exposures to these suspected chemicals and overlook or deemphasize known carcinogens, where exposure reduction would result in "known" reductions in human health effects.

Without explicit considerations of the uncertainty associated with any decision, a determination of confidence in real terms for a decision or a comparison with others is not possible. Uncertainty and confidence are inversely related. Without knowing the level of uncertainty one can not determine the level of confidence. Therefore the "reasonable confidence" approach in risk estimation may lead to "unreasonable" confidence in any given estimate or in the effectiveness of any action. The use of quantitative uncertainty analysis would greatly improve the overall confidence in decisions and selected actions.

Implications of Plausible Upper Bound Point Estimates

In addition to the use of uncertainty factors the use of a conservative approach where upper bound estimates are used can also seriously reduce the quality of the final risk estimate. The cascading of upper bound estimates disguises the underlying uncertainties. Risk assessment in the regulatory process is an ordered evaluation which systematically incorporates data from several sources, possessing upper bound estimates on their own (unless alternative information is available).

However, the use of upper bound estimates aggregately, piling one atop the other can lead to an extreme upper bound estimate of risk. In Superfund this is especially of concern as standards from other programs utilizing conservative estimates of their own are applied out of context in the Superfund program. This method of "extreme prudence" can grossly overestimate the risk by an unknown degree, supplies too narrow an estimate and "can only lead to excessive fear, a misalignment of priorities and misallocation of resources."(Diesler, 1987) While the use of upper bound estimates may be a legitimate policy position, these policies should be explicitly stated. "Misrepresenting the upper bound estimate as the risk to individuals ...blurs the distinction between significant and trivial risks and makes it difficult for regulators to set priorities or to allocate resources."(Scroggin and Kerester, 1988)

The cascading of upper bound estimates is potentially a great source for confusion and poor management decisions, in that evaluating or comparing actions in "real terms" is not possible. Because no determination of actual risk reduction can be made, no inference of incrementalism can be determined, either in cost or in evaluation of options.

Use of Animal Data

There are many who feel as though there is no validity to the extrapolation of animal studies to the human condition, that the differences between humans and other lower species are too great to justify such usage. While it is acknowledged that differences do exist, there are advantages to the use of such evidence of effect. Major differences between humans and other mammals include variability in: specific metabolism, rates of absorption (gastro-intestinal, skin and respiratory), GI tract microflora (affecting intestinal absorption, metabolism and transformation), lower skin permeability in humans, and qualitative respiratory differences (e.g., rats are obligate nose-breathers).

The existence of animal data is a valuable source and should be utilized. However, greater attention should be paid to how this data is integrated and guidelines established to represent the uncertainities in the final estimate. The existing method of using a uniform uncertainty factor (of 10X) is not sufficient and the uncertainty factors should be developed more on a case by case basis. Animal data should not be used alone, but only when it is supported by either physiological pharmacokinetic modeling and/or structure activity relationships.

Existing Health Databases

The mechanism of assuring consistency is the two major databases used in risk assessment, IRIS and PHRED. These

databases are available to and used by everyone associated with Superfund risk assessment. While these provide consistency they may also be a source of error or misrepresentation.(Category 1 and 4) There have been criticisms that the quality control for the input of data into these health effects databases has been deficient, and that unverified studies and data are being used as the basis of decisions (Categories 1 and 4). Therefore unverified and potentially inaccurate data may further cloud any real estimate.

Criticisms have also been levied that the information presented in the databases is not complete and does not include all studies. Negative study results are deemphasized as a single positive result (e.g., NOAEL) is frequently presented alone with little mention of other results. Given the reliance on the identification of a NOAEL from reported study results and the use in establishing a standard, inaccurate or misrepresented NOAEL results may lead to incorrect standards. When a NOAEL is identified, other information from studies showing dose response effects at slightly higher exposures may not be considered. A criticism has been made that if the identified NOAEL study results are an artifact then the resultant value may be possibly unprotective.(Category 1)

Looking at all the data may give a different value. That is, the dose response relationships indicated by other studies may actual predict a higher NOAEL than was observed from the single study identifying the NOAEL. Therefore a mechanism should be developed to incorporate all available data, possibly by some weighting mechanism.

Apportionment/Fractionation

The SPHEM requires that target concentrations be developed for remedial action development. In instances where there are multiple chemicals present, the target concentrations for each are reduced to compensate for other chemicals present. This apportionment is often across the board and uniform (if there are 5 chemicals present each would be reduced by a factor of 5 below its own target). Considering different levels of uncertainty in each of these compounds and/or exposure route, this crude approach may have little effect on overall actual reduction in the number of lives lost or on the confidence that an overall goal will be achieved. More effort should be made to base decisions for apportionment on real terms, by accounting for differences between the uncertainties associated with specific chemical standards, health data and/or exposure routes.

Consistency of Application

Risk Assessment Practices

Risk assessment as defined by EPA procedures is designed to optimize consistency of application. This is assured by the use of uniform models, accepted assumptions, and consistent databases. Despite its design for maximum consistency there have been several concerns raised as to whether risk assessments carried out under Superfund have been consistent or the object of manipulation.(Categories 1-4) Some feel as though there is a broad range in the quality of risk assessments, with major differences occurring between PRP-led and Fund-led sites. This is based on observations of those involved in the program. Further study is being undertaken to determine if the claims are substantiated. (Category 1)

Many of these discrepancies are felt to be due to liberties taken with the range of values for acceptable assumptions particularly in the exposure assessment. Extreme values of suggested ranges are frequently chosen when it is in a given party's best interest. To address this observed lack of consistency of quality the EPA is undergoing a program of quality control on risk assessments received. Also further guidance will be developed giving further direction on the use of these values and in general reducing the range of accepted values and requiring justification for deviating from that range.

De minimus vs. de manifestis risks

The experience of risk management and government regulation has been consistent. Governments are not responsible for regulating all risks, and must make choices as to where to allocate regulatory resources and effort. Travis et al (1987), reviewed a series of past governmental regulatory decisions (EPA, FDA, TSCA, etc...) concerning risks and have revealed patterns from past government actions. Two thresholds in the decision process were identified and defined as de minimus and de manifestis risks.

De minimus risks represent a risk level below which is deemed to be acceptable or sufficiently small that action is considered unwarranted. According to Travis, past decisons have indicated that this is a risk of 10^{-6} for excess lifetime probability of death. De manifestis risks represent a level of risk above which is clearly unacceptable or sufficiently great that actions or interventions to reduce the risk are required. Travis, et al, have defined the implicit value for a de manifestis risk level as 4×10^{-3} (all substances with preregulatory risks greater than this value were regulated).

Cost was shown to be a contributing factor influencing decisions within the de minimus-de manifestis range, in that for some regulation actions taken the costs was less than or equal to $2 million per life saved, while for some actions not taken the costs per life saved were greater than $2 million. The protective range as defined by EPA in Superfund matters (10^{-7} to 10^{-4}, with 10^{-6} as the target) is conservatively within the de minimus and de manifestis range.

Therefore the flexibility provided in the use of such a range is consistent with the history of EPA regulatory actions. However, Superfund does depart from past experience when costs are considered. The implicit value of life of $2 million per life saved is not adhered to in the Superfund program in selecting remedial actions. Actual costs per life saved for remedial actions have frequently been shown to exceed this value by several orders of magnitude. (OPPE, 1987)

Demonstrability of Goals

With the uniform use of assumptions and models the resultant standards can be established as the goal. In the case of Superfund actions this goal (of protectiveness) has been defined as attaining ARARs, a hazard index of less than one, and a range of 10^{-7} to 10^{-4} for excess individual lifetime cancer risk. This goal can be used as a standard for comparison and a determination if the goals have been met. By definition all Superfund remedial actions must meet these goals and therefore goals are demonstrated or met.

However, when considering risk in "real terms" and the uncertainties associated with these risk estimates demonstrability of goals in real terms is not straightforward. Predicted risk values represent plausible upper bound estimates and may not be indicative of expected actual risks, so comparison with goals or risk reduction in real terms is not possible. Therefore in terms of the defined standard of goals demonstrability is straightforward and concrete but is not possible in real terms under present risk assessment methodologies.

F. ALTERNATIVES TO THE USE OF CONVENTIONAL RISK ASSESSMENT IN ESTABLISHING CLEANUP LEVELS

As shown in previous sections there are significant criticisms of the Superfund program. These are not limited to the risk assessment portion of the process, and may or may not be legitimate. Public confidence in Superfund is not optimal. The major criticisms can be summarized as inconsistency, lack of permanent cleanups, postponement or transferrance of risks, and a reliance on least costly solutions.

To improve the overall program and increase public confidence, the EPA is undertaking an evaluation of alternatives to the present process of establishing cleanup goals which relies heavily on conventional risk assessment. Through the process outlined in the Prologue the following major alternatives were identified and will be presented: concentration-based standards, best demonstrated available technology (BDAT), improved risk assessment, and benefit-cost analysis. Some of these alternatives have already undergone varying degrees of development within the EPA. An additional alternative which based cleanup levels on background or pristine levels was eliminated from consideration because it did not make environmental, technical, or economic sense (OTA, 1985).

Determinations of the extent to which cleanup is necessary at hazardous waste sites can be divided into two basic categories: absolute and relative. The absolute approach is based on the assumption that acceptable concentration of a hazardous substance in an environmental medium can be defined, independent of site conditions, at which no significant adverse effects to human health or the environment would be expected. The standard or absolute value adopted would then define "clean". Alternatively, the relative approach would allow the determination of clean to be made on a situation or site specific basis. Such a determination would consider and may be driven by site-specific conditions and pathways, technology, costs, risk/benefit analysis or percent reduction. "Clean" in this case is defined as the most acceptable option in a particular decision problem(Brown, 1986).

The relative approach provides the opportunity and framework to incorporate the multi-disciplinary factors impinging on such a decision. The benefit of the absolute approach is that once a standard is adopted it becomes simple and non-controversial, more defensible, provides a communication link between levels of skill and knowledge, appears objective and scientific, and simplifies the problem (Brown, 1986).

I. Absolute: Uniform Concentration Based Standards:

Description

A common criticism of the Superfund program and past remediation actions has been a perceived inconsistency of outcome on the part of the public. It has been difficult for the general public to comprehend that a treatment technology or residual concentration considered safe and protective at one location would not be considered safe at another location. This variation in safe residual concentrations and use of technology has caused great confusion, vocal criticism, and a public mistrust of the Agency's management of the program. SARA specifies that the Agency develop consistent national cleanup standards for any residual contamination left on site, which in the public's view implies concentration standards which define "clean." The alternative which most directly addresses the perceived inconsistency in outcomes is the development of uniform concentration-based standards which define "clean".

Uniform concentration-based standards would be developed for the chemicals encountered at CERCLA sites establishing what is considered "clean". Standards would be developed on a chemical-specific and medium-specific (e.g., groundwater, soil, etc...) basis and based on generic fate, transport and exposure scenarios. Clean would be defined as the point where all contaminants are at or below the established concentration standards. This approach would provide complete uniformity among sites in terms of residual concentration.

Required in this generic exposure scenario is attention to site-specific subsurface considerations such as soil type, which may effect which generic fate and transport scenario is selected. A limit for depth of treatment would have to be defined. This limit would be based on assumptions for what depth would be expected to remain undisturbed under normal conditions for future site development, under normal construction activities (e.g., the depth of excavation required for a foundation in a home with a full basement for a predicted residential area).

A waiver provision would be made available when specific site conditions, such as fractured bedrock, prevent attainment of the standard. This waiver would apply if the standards can be shown to be unattainable for technical, economic or health reasons (if greater risks would result from the implementation of remedial activities as indicated by a generic risk assessment compared with a baseline risk evaluation). The burden of proof would be on the Agency and documented.

Concentration standards can be established using one or more of the following methods:

(1.) *risk-based*: Existing health and toxicological data as well as conventional risk assessment techniques (using a generic fate, transport, and exposure scenario), would be utilized to establish interim standards for substances found at contaminated sites. This would be carried out in the same way that target concentrations are presently calculated. Standards would be developed to result in uniform risk levels based on a generic fate, transport, and exposure scenarios. These concentrations would be defined as protective on a chemical specific basis. The target risk level for the standards for any contaminant would be within the 10^{-4} to 10^{-7} excess lifetime cancer risk range with 10^{-6} as the point of departure, as is defined in the existing Superfund process. The point of departure would be deviated from only in special technical or economic circumstances, such as when the limit of detection is greater than the predicted target concentration.

The uniform generic exposure scenario would be developed for residential uses (considered the cleanest goal) and the standards based on these values. Standards developed for a residential setting would be assumed to be conservative. An industrial use factor would be multiplied by the residential standard to determine the allowable concentrations in an industrial area.

(2.) *technology-based*: Experience from previous cleanups, academic and engineering research, and other technical experience treating similar wastes in industry, would be develop the standards. As with all technology-based standards, regular review of technology and standards would be required.

Standards may be based on one or more of the following:

i.) Residual concentrations would depend on what is considered technically achieveable given the state-of-the-art technology for treatment. State-of-the-art determinations would be based primarily on engineering research and the application of technology from other related uses (within industry). This would promote treatment and the implementation of new treatment technologies.

ii.) Technically feasible remedial actions would serve as the basis for the standard. Technologies which are in use are considered feasible in the absence of contrary information.

iii.) Standards may be set at the existing level of detection for those chemicals where the detection level is higher than the corresponding 10^{-4} excess individual risk.

Evaluation

Completeness of Conception

The concentration-based standards approach does not afford the opportunity to represent a complete model. Assumptions and simplifications must be made to develop a generalizable standard. The process of generalizing requires a simplification in the problem definition. Therefore under this alternative it is likely that many important factors will be left out of consideration. Some of the concerns not represented in a concentration-based standards approach include a measure of overall protectiveness, additivity and synergism, direct technical and economic review, site-specific characteristics, technological development (for health-based standards), and balancing the Fund.

For health-based standards, concerns of risk are believed to be indirectly incorporated into the standard. Likewise for BDAT-based standards, technical and economic feasibility is believed to be included. Only the state-of-the-art technology approach considers or promotes technological advancement.

While not documented there appears to be a general acceptance of standards by the general public(Category 3). This standards-based alternative would exploit this general acceptance of standards and most likely expedite acceptance and implementation. The numbers (standards) would likely assume personalities of their own, and as risk is never mentioned explicitly this may eliminate the element of risk communication and its associated problems.

The standards alternative would reduce the perceived misuse of cost-effectiveness (to choose the least costly) by explicitly defining clean and then choosing the least costly solution meeting that goal.

Rigor of Thought

Owing to the simplified problem definition required for a standards-based approach, the level of detail afforded to various elements of the Superfund problem is minimal. The absolute standards approach does not have the ability to balance the many interests and factors which may influence or be affected by an environmental release of hazardous contaminants. This alternative lacks flexibility, making consideration of new concerns or goals difficult.

There are many practical constraints to the development of such a large number of standards. The resources required to develop standards for the vast array of chemicals found within Superfund would likely exceed the capacity of the EPA. This is indicated by the significant expense and time already required to develop the relatively few existing standards.

Uncertainty in Health Estimates

The standards alternative does not allow consideration of, and may actually mask, uncertainty in the decision making process, which may lead to a misallocation of priorities and resources. Health-based standards would utilize conventional risk assessment. The uncertainties and limitiations associated with risk assessment would not only still be present but compounded by new uncertainties introduced by the simplified problem definition (e.g., generic fate and transport assumptions). Also, by adopting single value standards the uncertainty inherent in the development of those numbers is removed from further observation and evaluation.

Technology Definition

For technology-based standards the problem lies in defining technology as feasible, BDAT, or state of the art. Previous experience is typically used to defined "demonstrably available". The test of prior usage defines or gives credence to its feasibility. However, feasibility is a complex parameter and is influenced by many factors. The feasibility of a technology is related to the societal will to promote its development, achieve specified goals, and commit resources, both personnel and financial. Therefore the narrow definition of feasibility defined by prior usage may limit the technological options available.

An associated consideration in the development and use of new technologies is performance uncertainties. Technologies considered state of the art can achieve better performance than exisiting technology but may be unreliable. Therefore to establish standards based on theoretical or bench-scale performance may not be prudent. The standards would have to set at a level that is reasonbly achievable in common practice. The use of proven technologies rather than state-of-the-art would minimize performance uncertainties but might result in lower defined achievable levels.

Adequacy of Actions

On the basis of risk and cost, both excessive and inadequate actions would be expected when compared to the existing measure of protectiveness. The use of standards does not allow the balancing of other issues. An example of an

excessive cleanup can be shown in the depth of contamination considerations. The 8-10 foot depth is used as a standard to protect against exposure from future excavation. This would require the removal and disposal of large quantities of dirt at great expense. An inadequate cleanup might be one where there are a large number of contaminants or synergisitic effects. These additive and synergistic effects might result in the remedial action exceeding the protective risk range.

Excessive cleanups, while they may be good in the sense of providing greater human health and environmental protection, divert resources from other potential cleanups. This would be undesireable and an example of an opportunity cost and Fund balancing. Insufficient cleanups have the impact of subjecting a population to unprotective levels of hazardous contaminants. The standards approach is inflexible and does not allow a balancing of tradeoffs to meet overall EPA goals.

This alternative essentially requires a clean closure by defining allowable residual "clean" levels. Experience has shown that clean closures are the exception rather than the rule. Should a waiver provision be required for a majority of actions this would reduce the attractiveness of this alternative, requiring redundant effort obtaining waivers. This may also reduce the overall confidence in program manangement similar to the existing problems with ARARs. The EPA has been criticized for ignoring ARARs without obtaining waivers, while the EPA has felt that it is unnecessary effort to continually obtain waivers for infeasible ARARs.

Total Risk Considerations

The experience of standards is inadequate to address multiple contaminants and multiple routes and so it is incapable of addressing total risks, either from a site, a medium, or a chemical. Such a total risk approach in standard development would require significant changes to be made in the philosophy and mechanics of standards development throughout the Agency, which is certainly not feasible in the short term. All standards which have been established without consideration of multiple contaminants (in essence all standards to date) would have to be reevaluated. Such an approach may also risk the possibility of rendering all previous standards developed by the agency unprotective and invalid.

Feasibility

A technology-based standard would be incapable of considering the implications of multiple compounds on feasibility and treatability. Standards are developed on a

chemical specific basis utilizing a given technology, ignoring the influence of the presence of other compounds. Evidence for feasibility or treatability for a single compound may no longer be relevant to mixtures.

Consistency of Application

The standards-based alternative assures a consistency based on the residual concentrations of contaminants only. The standards define clean and are applied uniformly nationally. There is no assurance or that there would be consistency among sites on the basis of other factors. Costs and residual risks would vary greatly between sites depending on the extent of contamination and the actual exposures, respectively.

The use of standards as cleanup goals does eliminate the criticisms of risk assessment manipulation. Presently liberal exposure assumptions are used in setting target concentrations for an industrial setting. These assumptions for use in an industrial setting are believed to be the object of manipulation(Category 1). The use of a residential standard multiplied by a uniform industrial factor eliminates this source of manipulation and variation.

'Demonstrability of Goals

This alternative offers excellent demonstration of goals if as narrowly defined as the residual contamination concentration standards. However, this alternative is deficient in demonstrating or achieving goals represented in the other evaluation criteria. There is no mechanism to insure or even demonstrate protectiveness beyond the limited definition of the standards. This alternative is severely deficient in allowing a balancing of the criteria and other issues.

II. Relative

(a.) Designated Treatment Technology

Description

To promote treatment a matrix of technologies believed to be the Best Demonstrated Available Technology (BDAT), both economically and technically acceptable, would be developed. This selection matrix would be based on concentration, medium, and site characteristics (by category). On one axis would be concentration of a compound or class of compounds broken down into several concentration ranges, on the other axis would be the medium or site descriptor (e.g., soil type). Within the cells of this selection matrix the designated BDAT for those conditions would be listed. BDAT is not limited solely to treatment but includes source and institutional controls depending on contaminant concentrations.

Indicator chemicals would be chosen much as in the existing method outlined in SPHEM and would represent those chemicals posing the greatest risk. These indicator chemicals would then be used in the matrix to select the appropriate BDAT for the given concentration range. It is noted that while multiple chemicals would be present the BDAT is chosen for only indicator chemicals. While the technology is chosen for a select group of chemicals it is assumed that the chosen BDAT would also reduce the concentration of other similar chemicals present. The resultant levels of cleanup would be assumed to be adequate and protective, reduced to the lowest that is technically feasible. Performance and design standards associated with each technology will help to ensure that a technology functions as anticipated.

BDATs would be chosen based on existing health goals, past experience of use in industry, remedial actions already undertaken, and existing technical knowledge. Past usage of a technology would indicate a level of economic and technical feasibility. Experience of treatment in an industrial setting would be used to augment the limited experience with specific Superfund cleanups. Development of standards (i.e., % reduction) should consider the existence of chemical-specific health goals and ARARs such that the specified level of treatment would approach them as best as possible, but would not be required to do so.

Treatment technologies would be reviewed on a regular basis and the selection matrix would be changed to reflect technological improvements. A formal review/ approval mechanism would be developed to expedite the implementation of innovative technologies in periods between the regular revisions of the selection matrix. An application could be made to a formal review board (consisting of technical and

economic experts) who would review it on the basis of its technical merits, the results of required (bench through pilot) tests, and costs. The board would then make a formal decision. Acceptance of an innovative technology by the board would result in its listing as an acceptable alternative for technologies designated within the selection matrix for given concentration ranges within a particular medium. The formal listing would then become final during the next regular revision of the selection matrix.

One of the formal goals of Superfund is to promote treatment or permanent solutions over temporary measures. The EPA has developed the Superfund Innovative Technology Evaluation Program (SITE) to promote such activities. This program should be expanded to take an active role in the listing of innovative technologies. While the technologies under review are not proven and therefore not designated BDATs, their development should still be encouraged.

Superfund sites typically have a mixture of many chemicals of differring classes of compounds. Therefore it is likely that different indicator chemicals may actually result in the selection of differing or opposing treatment strategies. Treatment for one class of compounds may not be suitable for treatment of others (e.g., treatment of organic compounds may not affect the concentrations of metals present). Criteria must be developed for the selection of treatment process trains, either in series or parallel, to maximize the effectiveness of overall treatment and minimize residual levels.

A waiver program must also be developed to allow for unique site specific conditions which do not allow for treatment, such as fractured bedrock. Provisions, standards, and guidances are required to identify these conditions as they arise and to develop alternative remedial strategies.

Performance standards are to be developed based on statistical levels of treatability from a technology performance data base such that processes would be expected to meet the levels 99% of the time. Design levels are likely to be lower than that maximum achievable, due to the need to design treatment systems which will consistently achieve the standards. Concentrations are developed for individual contaminant levels, even if multiple contamininants are present in each chemical group.

The Office of Solid Waste and Emergency Response is developing an information data base on the experiences of the selection and performance of specific technologies in the Superfund program. This information base will summarize the experiences of the program to date, determine trends, and yield guidelines or precedent for further actions. This database is not being developed for the purpose of

designating technologies as proposed in this alternative; rather, it is being developed to augment the present process of remedial action selection, serve as guidance, facilitate the process, improve communication, and reduce information needs (eliminating the practice of "reinventing the wheel" at each site).

Initial trends or guidances suggest that if the contamination is significantly higher (e.g., 3-4x) than health-based levels then treatment is required or suggested, and if contamination is close to health-based levels then containment is the likely remedy of choice (Category 1).

Evaluation

Completeness of Conception

This alternative offers a fairly broad consideration of the contributing issues in a Superfund action. The selection matrix represents much of what would be required under the goal of the existing RI/FS process, to choose the best treatment option (remedial action). By utilizing a designated technology approach the RI/FS process is greatly acclerated, thereby expediting the selection of a remedy and the initiation of the design and procurement process.

Under this alternative, permanent treatment solutions would be emphasized, consistency of technology application is assured, and a mechanism to insure performance is also included. By using proven feasible technologies economic Fund balancing conditions are incorporated. Through the promotion of treatment, future concerns and impacts are reduced, although they may be difficult to define or quantify .

Technology and its performance is not immune from uncertainty. Elements of technology which have associated uncertainties include: failure rates, costs, and temporal extrapolation (time to treat, plant lifetime), size (bench scale to plant scale), and space (geology, hydrogeology, ecology). A BDAT approach merely exchanges the uncertainties associated with risk assessment for those associated with technology and its performance.

The discipline of engineering is more advanced and developed than that of risk assessment, and it has a more tangible experience base to draw from. In general, the confidence in technology is greater than that in risk assessment. However, without a complete understanding of the overall uncertainties and their implications it is not possible to determine conclusively which choice is more advantageous.

The use of BDAT would result in public acceptance by those for whom BDAT implies the cleanest achievable, protective,

and that all that is possible is being done, ("the best that we can do"). Futher, the use of BDAT circumvents risk communication and its related problems. By eliminating the explicit discussion of risk, the question of or attention to a specific measure of protectiveness is minimized, regardless of whether remedies are protective.

The use of BDAT would promote environmental protectiveness in cases of limited human exposure. Existing methodology tends to emphasize human exposures and human risks resulting in situations where some cleanups may be compromised due to the scarcity or unlikelihood of human exposures. The BDAT approach is linked to concentration within a medium, the ability to reduce that contamination, and is not limited to human exposure.

Although while it will be covered later under the demonstrability section, it should be mentioned here that the BDAT approach does not insure a protective remedy. While treatment would be carried out, there is no assurances of protectiveness when considering additivity, multiple routes of exposure, and multiple chemicals.

Rigor of Thought

The criteria for feasibility may not be applicable to a multiple contaminant situation. While a single BDAT itself is technically and economically feasible, it does not neccesarily hold true when the technology is required to be used in series with several other processes or used for several incompatible compounds. When several BDATs are used in series the overall costs may prove excessive and no longer feasible and/or the resultant treatment performance may not be as predicted when used alone due to interferences from other processes or chemicals. Therefore, standards or guidances are needed to account for the design or use of BDATs in series and for establishing upper-bound thresholds above which the BDATs displayed in the selection matrix are no longer feasible (due to costs, technical interferences, or incompatibilities of selected BDATs).

The public acceptance as discussed above may in actuality be divided, as is the case with previous programs such as the Safe Drinking Water program. The acceptance of governmental standards (or the use of BDAT) is dependent upon people's feelings toward governmental authority and expertise and the desire of industry to create technologies in line with human needs. The general uninformed public may defer to the expertise of the government and its duty to protect the public health. However, those who are inherently skeptical of governmental authority may not be likely to accept this or any other alternative based entirely on it's merits. The acceptance by those individuals is generally accomplished in

the process of developing such an alternative, through their actual or perceived involvement.

The development of the selection matrix is conditional on the ability to categorize sites into a finite, manageable number of categories. Such a categorization may prove difficult and thereby impede the development and feasibility of such a selection matrix. A review of the literature and criticisms, as well as personal interviews, indicated that there is not much agreement as to the viability of carrying out such a categorization. Opinions from interviews seem to be evenly divided as to whether sites can be categorized in 10 categories or fewer. The success of the designated technology alternative may depend in part on the ability to have a workable number of categories within the matrix.

The designated technology (and the selection matrix) is linked to concentration and medium to establish zones of management within the project site. A threshold concentration exists for a medium above which is addressed in the selection matrix but below which is not included. Due to migration of contamination there may be instances where the spread of contamination may occupy an area greater than that designated as the project site. While the concentrations would be below those levels requiring attention there is still residual risk associated with those concentrations.

The magnitude of that residual risk is related to the extent of contamination and exposure. In some circumstances this residual risk may be significant. The designated technology alternative does not address the risk within the no-management zone and, therefore, could result in excessive risks to the public.

The potential exists in the BDAT alternative that unique subsurface geological and hydrological conditions may not be adequately addressed. Such conditions may preclude using a particular technology of choice unless there is an adequate waiver procedure. Even if the waiver provision exists, there is still a chance that inadequate site investigation or characterization may not reveal the unique circumstance requiring such a waiver. Without the recognition of the anomaly and use of the waiver, a technology may be selected which may be inappropriate, costly, and ineffective.

A general criticism of any BDAT approach is that it does not tend to promote the development of new and innovative technologies. Technical improvements tend to be directed at lowering costs for accepted levels of treatment as opposed to attaining further or greater levels of treatment. In theory BDAT is the technology which most closely approaches the health goal for a particular constituent at a reasonable cost at a particular time, and periodic review of technology

and its improvements should approach the health goal over time. However, the experience of other programs based on BDAT (e.g., the safe drinking water program) have shown that the trend of approaching the health goal is slower than originally anticipated at the time of inception of the program. Thus the improvement of technologies under the BDAT approach is only as strong as the commitment to review, including commitment of funds and personnel, and the economic incentives for such development.

Consistency of Application

Consistency in outcome would be readily achievable with universally applied accepted treatment technologies. Through the use of the matrix, technologies would be selected consistently thorughout the country. Likewise, since the technologies would be uniform, by implication the level of treatment also would be uniform throughout. In addition, as the early stages of the RI/FS process would be truncated, the potential for manipulation or bias (e.g., through risk assessment or costing assumptions) would be reduced. Therefore, it is likely that this, too, would promote further consistency.

Demonstrability of Goals

A designated BDAT approach requires treatment when feasible as part of any remedial action; therefore addresses and eliminates a major criticism of the Superfund program, that permanent treatment rather than temporary source control options are used too infrequently. Source control would still be a designated BDAT in some defined circumstances, but the option would be deemphasized overall.

While the treatment options presented are considered to be economically and technically achievable, protectiveness of human health and the environment has been implied but not assured, and no provisions are made for multiple contaminants or unusual exposure characteristics. Exposures, existing or potential, are not assessed, and therefore there is a probability that individuals or populations may be subjected to excessive risk or that environmental damage may occur.

By definition BDAT is the best demonstrated technology and implies both technical and economic feasibility. Cost-effectiveness is inherently incorporated into the process as the basis for technology selection. The use of proven economically feasible technologies would yield a cost effective program, if not in practice, at least in perception. BDAT perceptively promotes overall program efficiency by utilizing cost-effective measures and promoting permanent treatment (eliminating recurring cleanup costs from failed actions).

(b.). Improved Risk Assessment Alternative

The exisitng process of establishing cleanup levels relies heavily on conventional risk assessment. Many of the criticisms of the remedial selection process are rooted in the apprehensions of the public toward risk assessment. It can be inferred that the public's confidence in the effectiveness of the Superfund program is related to its confidence level in risk assessment. Overall improvements in and acceptance of risk assessment methodologies would likely result in greater public program acceptance.

As described in Section D, there are many recognized shortcomings of risk assessment methodologies, both on a theoretical level, and, more importantly, in its application. While the discipline of risk assessment is perceived as being in its infancy, there are many techniques available which have yet to be incorporated into the simplified conventional risk assessment paradigm. As stated in Section D, the conventional risk assessment approach is one which has been simplified to achieve consistency at the expense of completeness. However, modifying the paradigm through the incorporation of more sophisticated techniques for both the estimation of risk and how its presention in the decision process would result in a higher level of sophistication and confidence in the overall process.

Description

The goal of this alternative is not to improve the science on which risk assessment is based but to improve the use of existing techniques for calculating risk given existing knowledge and its presentation in the decision process. The EPA is presently pursuing a program to improve the science on which risk assessment is based. The Reducing Uncertainty in Risk Assessments (RURA) program focuses on improving the science of generating risk estimates and reducing overall uncertainty in those estimates. Areas targeted for research include structure activity relationships, physiological and pharmacokinetic modeling. While this is a worthwhile endeavor which will benefit any path the EPA chooses, it will not eliminate uncertainty altogether.

Decisions will still have to be made under conditions of uncertainty regardless of the successes of RURA. Merely reducing uncertainty or improving information does not always improve overall decisions. Early EPA decisions that established standards took months instead of years, involved few people, and were based on limited existing knowledge. It could be said that these decisions were based on what could be termed "a highly educated guess." Public acceptance of those decisions and subsequent (and more sophisticated) analysis has confirmed that those decisions were reasonable.(Yosie, 1987)

This alternative, therefore, is not directed at improving the science on which risk analysis is based but on improving risk management decisions by improving the manner in which risk estimates are derived and represented in the overall decision making process. This alternative will focus on two major elements: uncertainty and sensitivity analysis. The existing risk assessment guidelines and the RI/FS process recognize sensitivity and uncertainty analysis as important, but their analysis is carried out predominantly in a qualitative manner. Under this alternative, when possible, formal quantitiative techniques for sensitivity and uncertainty analysis would be required.

A major component of the alternative is to utilize all relevant models and data. Competing sets of inferences are available for drawing inferences (of risk) with no single set possessing complete dominance over its competitors. As a result it is never certain which set of experiences ought to be used in supporting the infrastructure of a decision (Crawford-Brown, Cothern, 1987). For example, EPA's standard practice of using a uniform set of models (e.g., linear multi-stage) limits the completeness of a risk estimate.

Using the results of all studies and models an estimation of risk can be developed which presents the range and likelihood of the estimates. The final representation of the risk may take several forms but would likely take the form of cumulative probability functions. These cumulative probability functions would be derived from probability density functions of the likelihood of each estimate (using a discrete set of parameter values and model) within the range.

The derivation of the cumulative probability function could be carried out by either elicitating expert judgement (as in Morgan et al, 1984), following a Bayesian method (as in Crawford-Brown, Cothern, 1987), or using stochastic or Monte Carlo modeling. The elicitation of expert judgement is a method of representing the subjective judgement of the scientific community involved in relevant research. The Bayesian approach provides a relative ranking of degree of support offered to possible inferences of risk based on both qualitative and quantitative information.

This prediction of risk as a range, instead of a specific point estimate more accurately represents the risk and provides greater information to the decision maker, hopefully leading to better decisions. An over simplified example of how these cumulative probability functions would appear and possible interpretations is shown in Appendix A. This method would allow a graphical review of the uncertainties in the overall risk estimate.

The EPA presently follows a similar procedure reviewing models, data, and overall risk estimates during SAB review for the development of standards. During that time the information is reviewed and balanced with other issues and concerns such as economics. That process may be applicable to a situation where a uniformly applied standard is to be established. In that instance the broader representation of risk is not necessary for its implementation. Superfund, however, is not an analogous situation because the decision similar to that made by the SAB needs to be made on a site-specific basis.

The presentation of risk as a cumulative probability allows decisions to made at known confidence levels which otherwise could not be made by aggregating plausible upper bound values. By using cumulative probability function one can select the confidence level for the overall risk estimate not just for the individual parameters, or compare costs to achieve overall confidence levels. For example, the protectiveness measure could not only be 10^{-7} to 10^{-4}, but an overall confidence in the estimate of risk of 95% for a 10^{-4} excess lifetime cancer risk.

The development of this expanded representation of risk could be carried out by individuals presently responsible for similar activities (such as CRAVE and the RfD workgroup). Once developed, this library of expanded presentations of risk would be incorporated into existing databases (IRIS and PHRED) and made available to those involved in risk assessments under Superfund.

Sensitivity analysis would be carried out in a formal manner for both the risk estimate and the decision model for remedial action selection. Sensitivity analysis determines the influence of individual parameters on the overall estimate. Those parameters which greatly effect the final estimate (sensitive parameters) are identified. Attention can then be focused on sensitive parameters (e.g., dose-reponse model, individual exposure paramaters, environmental transport assumptions, etc.) to improve the confidence in their values, thereby improving the confidence in the overall estimate. Confidence in the individual parameters can be accomplished by doing more in-depth analysis or possibly by obtaining more information. Value of information techniques can be used to determine the bounds on the cost of obtaining that additional information (Evans, Finkel). In this way the actual analysis can be structured to achieve the greatest confidence feasible.

Related to the improvements suggested under the risk assessment alternative is the field of decision analysis. Decision analysis provides a logical and systematic framework for structuring and analyzing problems.

Assumptions and the logic train used in an analysis are explicitly spelled out to allow a critical review of any decision and the determination of the influence of changing assumptions or goals on the outcome. Rodricks et al reviewed over 400 RODs and found risk information presented in less than 10%. Decison analysis is appropriate and helpful in solving complex problems which are characterized by: (1) many alternatives, (2)multiple, conflicting objectives, (multiattribute), (3)uncertainty, and/or (4)large magnitude of consequences. As the remedial selection process fits this description, decsion analysis would be appropriate.

The primary elements of the decision analysis paradigm are: (1) structuring the problem, (2) establishing goals, objectives, and measures of performance, (3) analyzing the alternatives with respect to the performance of the objectives, (4) setting up a multiattribute utility function, and (5) evaluation of the alternatives (including sensitivity analysis).

The techniques available under decision analysis would provide a clear way to improve the manner in which decisions are made and as a result the quality of decisions would improve. Work should be initiated immediately to develop a mechanistic framework to use as guidance in Superfund projects to aid in the formulation of the study and the decision process. As discussed earlier, many of the elements of decision analysis as applied to risk assessment (such as sensitivity analysis) would improve the RI/FS by helping to identify the sensitivity parameters for a given site, thereby allowing effort to be focused on those variables. The same approach can be applied to other components of the RI/FS process, site characterization, treatment technology performance, sampling, costs, and alternative selection.

Evaluation

Completeness of Conception

The improvements to risk assessment from this alternative would merely be an augmentation of an existing system. This alternative would utilize the process already in existence, and would benefit from the completeness of conception represented by the existing process. By improving an existing system one can use the accepted and proven existing framework and build on it. This should improve implementability and expedite initiation of this alternative.

Risk

This alternative provides as complete a representation of risk as possible. It is generally agreed that risk is not a single point estimate but is a distribution, or range of estimates or predicted outcomes based on applicable theories, models, and data. Risk has been defined as "probability *and* consequence" and may be further described as a triplet with the components: scenario identification or description, probability of scenario occurence, and the consequences or measure of damage for that given scenario (Kaplan and Garrick, 1981). Included in this definition is the concept that the range of models and data all have some validity and some probability of being correct. Ignoring any models and their results likewise disregards these potential scenarios from policy decisions. Therefore, the alternative presented here addressses, follows, and represents this risk paradigm.

Incorporation of Subjective Judgements

Multi-attribute decisions are not based entirely on logic and/or quantified information. Decisions are made by the subjective processing of information by an internal weighting scheme. Since both the risk assessment (because of the use of multiple components of information) and the selection of remedial actions are analogous to multi-attribute problems, the subjective element is important in these processes as well. This alternative yields a broader presentation of risk and allows the decision maker or the public to assess this full range of information based on their own subjective weights. This alternative augments the value and contribution of this subjective portion of decision process without forcing it into a formal model (with its associated introduced errors).

Rigor of Thought

Overall the existing remedial action selection process is believed to represent a fairly complete problem definition. This alternative increases that completeness by presenting a more thorough definition of risk, without removing any important elements of the process. This alternative also allows more rigor and depth in the decision process.

Uncertainty

The adequate treatment of technical uncertainty in quantitative policy studies or decisions is related to both the nature and the magnitude of the uncertainty. When uncertainty is small it may be appropriate to perform single value best estimate analysis followed by careful sensitivity analysis and to apply an uncertainty range to the final result. However, when uncertainty is great it is more

appropriate to use probability distributions to characterize the uncertainty. Uncertainty in these situations can then be propagated through the analysis either analytically or with stochastic simulation (Morgan et al, 1984).

This alternative could do either of these forms of analysis as the situation dictates. There is already some support for using best estimates in an analysis and applying an uncertainty factor to the final estimate (Category 4), but as indicated above even this approach is not appropriate in all circumstances. This alternative has the flexibility to choose the method most appropriate for a given circumstance. By not generalizing the model to be used, errors from that simplification are not introduced.

Sensitivity Analysis

The use of formal sensitivity analysis would reduce the manipulation of risk assessments, especially in the exposure assessment portion. For example, a sensitivity analysis would determine which of the exposure assumptions, if any (e.g., water consumption), are sensitive parameters, having an dramatic influence over the overall results (e.g., a 10% change in a parameter value that results in much greater than 10% change in the overall estimate). Once the sensitive parameters are identified, a closer look would be required and documentation would be necessary to support and justify the value chosen. For example, the standard assumption for water consumption of 2 liters/day/person may not be adequate in certain circumstances, and better information be required to determine if a value is available that is supported by evidence for a given site situation.

The use of sensitivity analysis can direct the emphasis of sampling and field studies toward sensitive parameters and deemphasize insensitive parameters. Potentially this would potentially reduce time and resources wasted spent unnecessarily to define parameters which have little influence on the overall estimate (i.e., when large changes in the value of a parameter have little or no effect on the overall estimate). The overall effect would optimize and streamline the process.

Prevention against bias

If all available data, models, and study results are used there is less opportunity for any particular study or model to dominate. Should one set of parameters be able to dominate, then the overall risk assessment is only as good as that set. There is always the chance that a model or study is flawed or biased and may have been selected in an attempt to manipulate the final results. By reducing the potential influence of bias the overall confidence in the assessments would likely be increased.

Comparisons/Commensurability

This expanded approach to the representation of risk allows a greater comparison among alternatives and sites and allows consideration of commensurability. Commensurability is the ability to compare different risk management decisions in terms of effectiveness. At present, since uncertainty is not expressed with risk estimates, detailed comparisons are not possible.

Comparisons are possible both between sites and within a site. The individual decision maker can tradeoff on confidence between given alternatives, possibly on the basis of cost per confidence level achieved or on overall weight of evidence. The best estimate (e.g., 50% confidence) would always be within the protective range. Overall confidence above that may be used as a tradeoff when compared with cost or feasibility. This assumes that an overall confidence level is not required to define protective but allowed to be a parameter to be traded off in balancing issues.

Limitations

In this alternative uncertainty is presented in greater detail, not eliminated entirely. Decisions still must be made in light of uncertainty and kept in perspective. It is possible that this representation of risk may assume its own meaning. Should these estimates take on their own meaning a reliance or exaggerated trust may be placed in this new and improved process causing one to overlook it's limitations. This exaggerated confidence may lead to carelessness in its applications, much as is the problem with the use of single estimates of risk. Therefore measures to be taken to insure the proper use of these estimates.

There may be a shortage of resources to carry out these improvements. The techniques already exist but most likely a shortage of personnel, both within and outside the agency, with the appropriate skill level to immediately implement this alternative. A delay for education or a phased implementation would be required. It should be noted that the delays necessary are still likely to be less than that required for the other alternatives. During the period of implementation there would likely be problems in the quality of these assessments and mechanisms of review and quality control should be stressed.

Consistency of Application

The consistency related to this alternative is one of process versus outcome. This alternative focuses on the consistent use of the tools and does not focus on the

outcome. If the tools are used consistently then the outcomes should be very similar, with differences due primarily to site-specific conditions. This alternative does not have the consistency of outcome that the previous alternatives possess. In other words, the consistency which is readily apparent in the others, such as residual concentration or chosen technology, is not present.

Measures will have to be taken so the methods are used properly. This will require educating those people responsible for preparing risk assessments, and using this information to make decisions about remedial actions. Furthermore, a mechanism for quality control to oversee the use of these tools would have to be developed. This should be acccomplished by utilizing the quality control mechanism being developed for existing risk assessments.

Demonstrability of Goals

The presentation of risk in a more broad sense allows comparison of alternatives and proposed actions; viewed in more "real terms" than is allowed presently using a single point estimate of risk. Greater confidence can be achieved in final risk determinations and that goals have been achieved.

The level of confidence at which protectiveness can be defined under this alternative can be chosen by the decision maker (e.g., a 95% confidence level that a 10^{-4} excess individual cancer risk is achieved). Tradeoffs can then be made for alternatives within the 10^{-7} to 10^{-4} range of individual excess cancer risk based on the other elements such as cost, feasibility, and acceptance.

The use of these cumulative probabilities allows greater confidence in the level of protection afforded and in the assessment of the effectiveness of risk reduction measures. Not only does this alternative yield a greater confidence in the final residual risk level estimates, it also provides a better check on protectiveness and efficiency. For example, a single point estimate of risk based on conservative values for parameters may be conservative, with a given probability of incidence below that point estimate. However, a given probability may also exist above the conservative estimate, and, more importantly, the uncertainty factors used in that estimate may not have been sufficient. By utilizing a cumulative probability of risk one can determine whether the single point estimate is protective or even within the range of plausible outcomes. A situation may occur through the cascading of plausible upper bound estimates such that the overall point estimate is well outside the range of predicted outcomes. Any effort to reduce the concentration on which that single point estimate is based would result in wasted resources as no "actual" reductions are realized.

While this alternative theoreticaly yields greater demonstration or confidence in the achievement of (protective) goals, there may be problems communicating it to the general public. A more sophisticated representation of risk requires a greater working knowledge on the part of the public. Expansion of EPAs risk communication efforts would be required. It is my opinion that this approach to risk presentation would actually improve risk communication and risk understanding on the part of the public. This enhanced understanding may also have positive implications for public understanding of other standard-setting procedures.

(c.) Benefit-Cost Analysis

Among the many criticisms aimed at the Superfund program and the cleanups it has undertaken is that large sums of money are being spent for cleanups of questionable integrity or for risk reductions of unknown magnitude. Conversely, a related argument frequently voiced by EPA critics is that the Superfund program frequently chooses the least costly as opposed to the most cost-effective alternative (as mandated by the law)(EDF, et al, 1988). Benefit-cost analysis can alleviate or address some of these problems.

A formal survey of selected EPA officials found that a majority (64.7%) support government use of benefit-cost analysis. Most of the support came from those involved in policy support and formulation, and opinion was equally divided among research and program areas (Rycroft et al, 1988). In addition, the committment to the use of benefit-cost analysis is apparent as indicated by Executive Order 12291. Executive Order 12291 was issued by the President on February 17, 1981. It requires that regulatory agencies prepare a regulatory impact analysis (RIA) for all major regulations. These RIAs must be reviewed by the Office of Management and Budget (OMB) prior to taking any regulatory action. In response to E.O. 12291 the EPA has developed a framework for benefit-cost analysis.

Description

The goal of the benefit-cost analysis alternative is the maximization of overall public health (reduction in risks) and environmental benefit for a given finite revenue base, or budget.

To address the concerns stated above and to promote more overall program efficiency this alternative uses benefit-cost analysis to determine the optimal "clean" solution. Specific site remedial action alternatives would undergo formal benefit-cost analysis during the planning and evaluation stages. Alternatives would be designed as dictated under the RI/FS process, and then selected on a strict benefit-cost basis. The cleanup would be measured in terms of maximizing benefits or utilities, and the most efficient solutions would be chosen. Those solutions would result in the greatest risk reduction per dollar spent, the greatest number of lives saved per dollar spent, the greatest extent of resources protected, or other measures of effectiveness.

Theoretically, an economically efficient action is one where the marginal cost equals the marginal benefit. Applied to the Superfund program one might say the equilibrium ("clean solution") is one where the costs of achieving greater

levels of clean exceed the benefit of doing so, and the cost of cleaning up less exceeds the benefits of allowing increased residual contamination. The costs and benefits which contribute to the equilibrium may be site specific and very different in different circumstances. "Clean" is where cleaner can not be achieved because of one cost and where dirtier cannot be allowed because of another cost.

In response to E.O. 12291 the EPA has developed the document Guidelines for Performing Regulatory Impact Analysis (RIA Guidelines), based in part on OMBs Interim Regulatory Impact Analysis Guidance but applied to an environmental setting. The EPA guidelines specifically address the techniques for benefit-cost analysis needed in the preparation of an RIA, including quantified and (when feasible) monetized benefits and costs and specific techniques drawn from past experience or developed in response to new requirements under E.O. 12291.

The existing benefit-cost analysis framework at the EPA would be used as a basis to develop a specific technique for application to the Superfund setting. The problems of Superfund and the nature of contamination and exposure at these sites differ from the targets of other regulations (e.g., ambient air and industries affecting air quality). Specific techniques must be developed to more accurately reflect the nature of the abandoned hazardous waste site.

Health effects (morbidity and mortality) are generally quantified by the risk assessment process outlined previously and are a quantitative cost or benefit. Morbidity should be valued by estimating direct costs and if feasible use willingness-to-pay to avert illness. Direct costs include medical costs, loss of work and earnings and impacts on future productivity. Mortality is valued indirectly or directly and uses "statistical lives saved" as the measure of comparison. Based on extrapolation of hazard pay figures the guidelines suggest that the value of a life should be within the range of $400,000 to $7,000,000.

While the application of a benefit-cost analysis may indicate an implicit assignment of a "value of a life", a good benefit-cost analysis can be carried out without making this distinction. Benefits (risk reduction) may be presented qualitatively or quantitatively allowing the decision-maker the opportunity to apply his or her own subjective value judgements. The benefit-cost analysis may actually take a risk-cost approach, where a comparison is made between alternatives based on apparent or predicted risks and the cost of achieving each respective level of risk.

When benefits cannot be easily quantified, a cost-effectiveness analysis may be used to evaluate alternatives.

The benefits of decreased pollution are the resulting improvements in human health, aesthetics, and reduction in damages to plants, animals and the environment in general. Benefits may be both quantitative and qualitative, be direct or indirect, and may be realized at the present or future.

Environmental effects are more broad-based and include effects on plants, animals, ecosystems, soils, and materials, damage, and adverse effects on recreational activities and aesthetics. There are four basic methods for valuing environmental effects: direct cost, travel cost, property value and contingent valuation. All lead to estimates of what groups are willing to pay to avoid damages of the pollutants in question and may measure health and nonhealth effects.(RIA Guidelines)

Evaluation

Completeness of Conception

In a benefit-cost analysis one is allowed to define the boundaries of the model and include all relevent concerns in a formalized structure. The model can be made to include the considerations of any party or interests and additional interests added when identified. Therefore the completeness of the model is dependent upon the representation of each individual concern. As a model has yet to be developed for the Superfund scenario it is not possible to comment on such completeness at the present.

Benefit-cost analysis provides the opportunity to represent a broad range of issues. Existing methodology used for remedy selection tends to emphasize human health effects. A proper benefit-cost analysis would also consider other consequences, such as environmental effects. It may be difficult to develop an agreeable framework for the adequate valuation of resources. Should this be accomplished, the benefit-cost approach would improve consideration of environmental and natural resource concerns.

Benefit-cost analysis can prove useful in establishing priorities and decision making within the de minimus - de maximus risk range. For risks outside of this range the actions required are quite clear. However, within the range the decision is often ambiguous. Benefit-cost analysis can be one method of clarifying the decision or better defining the important parameters.

Opportunity Costs

This alternative allows the consideration of opportunity costs, which have significant implications on the management of the Superfund program. Opportunity costs are those alternative actions (benefits) which are left unrealized by

the selection of and dedication of resources to any one particular action. In terms applied to the Superfund program, by undertaking a cleanup at one site (or to a given level), one is excluded from using those resources on another site and another cleanup. It is widely held that the present remedy selection process does not adequately address this issue, which is important to the balancing of the Fund.(Categories 1-4)

Through the explicit consideration of opportunity costs, benefit-cost analysis provides a mechanism of checks and balances to insure that excessive funds are not being spent on frivilous or marginal cleanups. Implicit in any benefit-cost analysis is the representation or quantification of benefits. If benefits are non-exsistent or negligible the expenditure of large amounts of funds for a particular remedy would be unjustified, given that those funds may be used elsewhere to realize actual or larger quantifiable benefits.

Use of ARARs

Benefit-cost analysis provides a mechanism to review conflicting mandates on which standards are developed. The Superfund program requires the attainment or exceeding of ARARs. The foundation of these ARARs is specific to the program for which it is intended. The goals or the basis on which standards are developed may not be applicable to a Superfund action. For example, consider the use of MCLs in establishing cleanup goals. The MCLs are established based on existing economic and technical feasibility, not only on health goals. The economic and technical feasibility for any standard is highly influenced by the medium and technology to which it is applied. Therefore benefit-cost analysis would allow the review of goals or standards in the context of Superfund.

For example, in the case of MCLs the standard applies to end of the pipe concentrations in commercial drinking water supply systems (>25 connections). This target (water treatment and distribution) has an inherent economy of scales for treatment technologies. These economies of scale may not be present in a Superfund remedial action. The costs to clean to such a standard may prove excessive. Conversely, technical feasibility may be influenced by the infrastructure, which itself may be a source of contamination (e.g., lead in drinking water), and lower levels may be achievable in a Superfund context.

Alternatives under benefit-cost analysis are not rigidly tied to the standards of other programs. Reviews of site or action specific relevant information determine treatment

levels which is efficient and feasible. The benefit-cost analysis clearly presents a mechanism for justifying ARARs or for obtaining waivers from the ARAR requirement.

Incorporation of Non-market Considerations

The value of perceived fear or of stewardship may explain the willingness of the public to dedicate the resources to Superfund cleanups. While the benefit-cost analysis framework would not be sufficient to use these as variables it does provide a framework for approximating their values and provides this information to the decision maker. If these intangibles can be accounted for the decisions made would more closely approximate the public will.

Temporal Considerations

Benefit-cost analysis provides a mechanism to review the temporal aspects of impacts, benefits, and costs. There have been criticisms levied that the Superfund is inadequately addressing temporal concerns, especially generational considerations. Some feel as though Superfund represents an injustice, that one generation is bearing the costs to treat the problems of a previous generation while yet a future generation is reaping the benefits. It is felt that the present generation is bearing all the costs without realizing any benefit.(Category 4) This alternative allows consideration of these concerns, either qualitatively or quantitatively. While the actual quantitative techniques may be limited and sensitive to assumptions made on discount rates, at a minimum qualitative information would be presented to the decision maker.

This alternative would provide a more rigorous framework to investigate and quantify future potential costs in review of alternatives. Institutional controls and capping have been criticized as being chosen as the least costly alternatives and ineffective (Category 3). While consideration of remedy failure and future costs are required under the RI/FS process, a more detailed explicit analysis would be required in this alternative. This more detailed analysis may have an overall impact on the usage of institutional controls, either justifying their usage, recommend improvements, and/or require greater committment of resources or funds and insurances for their success. At this time it is not readliy apparent what the overall trend would be, greater reliance on permanent remedies based on future costs or on improved institutional controls, but the mechanism for this evaluation is provided for under this alternative.

An additional temporal concern is that benefit-cost analysis allows for assignment of priorities and scheduling of remedial actions. This clear representation of actions would likely improve program development and accountability.

Rigor of Thought

The use of benefit-cost analysis is a fairly straightforward means of analysis for those actions which have readily quantifiable costs and benefits. However, when actions have consequences which are not easily quantified in economic terms or some overall common denominator the appopriateness of such an analysis is not guaranteed.

Incorporation of Non-market Concerns

The central component of any benefit-cost analysis is the definition of benefits and costs. For market goods this is a straightforward operation. However, for non-market goods this definition becomes more difficult. Superfund deals primarily with non-market concerns; the value of risk, perceived risk or fear, natural resources, the aesthetics of a clean environment, and stewardship for future generations. It is felt by some within affected industries and even within EPA program administration that the focus of attention on and expenditures by the Superfund program are not justified. The risks associated with a limited number of uncontrolled releases is felt to be minimal when compared to other universal or ubiquitous contaminants.(Categories 1,2, and 4) One possible explanation is that the non-market concerns must be of significant value for the public to value the Superfund program so highly.

The success of any benefit-cost analysis alternative is dependant on whether non-market concerns associated with Superfund sites can be incorporated into the process. Likewise, agreement must be reached on the assumptions used for market or monetized benefits. The difficulty of definition and quantification of benefits from environmental resources is a severe limitation to viability of a benefit-cost alternative.

The criticism that extensive funds are being spent on questionable or non-existent benefits, does not prove the non-existence of benefits, but does support their elusiveness or inability for quantification. Because of the inability for accurate quantification of benefits, it is felt this would result in an undervaluation of such benefits, lead to inadequate remedial actions, and deficient protection of the human health and the environment.

Value of a Life

Regardless of whether the benefit-cost analysis framework developed has an explicit expression of a value of a life in it's analysis, implicit inferences for such a value can be made. The quantification of a value of a life has many negative implications. If it were to become an accepted

element of the remedy selection process, it would be unacceptable politically and legally indefensible. The use of a value of a life would also have repercussions on other regulatory programs concerning consistency.

Limitations of Benefit-Cost Analysis

Benefit-cost analysis has its own general limitations. Many techniques for non-market valuation are available and results can vary greatly depending on the method used. These technique are also sensitive to bias. Future benefits are discounted from the future value to a value in present currency (the present value). The present value calculations are based on the discount rate, chosen as the predicted change in the future value of money. The discount rate and other assumptions can be contentious. The discount rate is a common point of manipulation in such analysis.

Benefit-cost analysis has it's own uncertainties. Should risk assessments continue to be used as a parameter in such an analysis, the inherent uncertainties in benefit-cost analysis would compound those already existing in risk assessment and other stages of the RI/FS process. Under a benefit-cost analysis alternative decisions would likely have to be made under even greater levels of uncertainty.

Consistency of Application

The benefit-cost analysis alternative would provide a consistency in process but not outcome (as related to residual concentrations). By utilizing a uniform framework for analysis and assumptions differences between actions would theoretically be based on site-specific conditions.

By utilizing a uniform framework of analysis and assumptions, it is hoped that socioeconomic, regional or geographic bias would be minimized. While a minimum of 10^{-4} excess individual lifetime risk would be met there would tend to be variance in the residual risk level between 10^{-7} and 10^{-4} excess risk range for remedial actions. The residual risk would likely be sensitive to the potentially exposed population, since cases avoided would be an important parameter. This would lead to a tendency to bias remedies to areas of greater population exposure, urban v. rural. This presents an issue of equity where rural areas may tend to be neglected in favor of the urban setting.

Conversely, if the framework of analysis decided upon includes the use of value of life in it's decision, consistency of protectiveness over socioeconomic classes would be assured. The benefit-cost analysis will provide consistent valuations only from the perspective on which they are based. In other words, for the value of life, if they are based on individual risk, the value of life would

be consistent across individuals from the perspective of the individual but would result in higher costs per death avoided in rural areas. Conversely, if cost per death avoided is the basis of calculation the cost would be consistent but the value of an individual life from the individuals perspective would be valued higher in the urban setting.

Demonstrability of Goals

Once a framework for analysis is decided upon and implemented overall program efficiency would be promoted through uniform analysis. Remedies would be selected on the basis of a balancing of program objectives, effectiveness and the overall fund. Utilizing a common set of assumptions and economic criteria, benefits would be maximized for Superfund as a whole given a finite base of financial resources. The explicit use and consideration of opportunity costs allows a mechanism of assuring maximized benefits.

However, due to the dominance of non-market elements in Superfund, if they are not adequately represented, a strict benefit-cost analysis would likely lead to cleanups and actions at a lower level of treatment than would be expected under the existing RI/FS methodology.

G. EVALUATION OF ALTERNATIVES

The nine evaluation criteria outlined in the RI/FS guidnace document are to be used along with the four presented in the paper (completeness of conception, rigor of thought, consistency, and demonstrability) as the dominant measure of evaluating the alternatives. As the nine evaluation criteria are the stated goals of the Superfund program and used to measure alternatives in remedy selection (and establishing cleanup goals) it is logical they should be used to evaluate the alternatives for how cleanup levels are established. In addition the four introduced criteria provide a broad but thorough approach to evaluating an alternative. In addition, two other criteria will be covered, program efficiency and equity.

Program efficiency concerns itself with the overall program of Superfund given constraints of a finite financial resource base (the Fund) and does not necessarily focus on decisions made on a site-specific basis. There has long been a debate within the EPA as to the focus of the program. Should emphasis be placed on maximizing the overall risk reduction (clean many sites well) or should sites addressed be cleaned to the maximum extent practicable (clean fewer sites better). The concept of program efficiency is meant to address this issue of the overall program view, evaluating the overall total consequences without compromising the cleanups at any particular site. The program efficiency parameter can be thought of as that which balances all the nine criteria to achieve an optimum.

Equity is concerned with the impacts on populations. Equity concerns are whether treatment is consistent across geographical or population sub-groups, the most common of these being economic status or class and population density (rural vs. urban).

The author carried out a subjective analysis or summary of the alternatives in relation to the above criteria. Results are presented in Tables G-1 to G-4. The alternatives were then ranked subjectively and a summary table of those results is presented (Table G-5).

Table G-5 contains only subjective rankings and does not imply distinct quantitative advantages for any one alternative. The rankings are relative within a given criteria and overall absolute judgements should not be inferred (i.e., the number of + does not indicate any magnitude of difference). An absolute evaluation will be attempted in a subsequent paper.

The table evaluates individual alternatives. Individual alternatives are considered as those mechanisms of establishing cleanup levels and are based solely on the

merits associated with that alternative. It is possible to develop alternatives which are combinations of the proposed alternatives (e.g., BDAT with risk assessment as a protectiveness measure). Should combinations be used, the ranking of that alternative to the others would assume the highest ranking associated with either of it's components.

A quantitiative framework similar to decision analysis could be developed to analyze the alternatives. Goals could be defined as could measures of performance (achieving those goals), the alternatives could be characterized in terms of these performance measures for each of the defined goals. An overall multiattribute utility function could be developed and used to quantify, rank, and weigh the alternatives. The measures of analysis could be based on levels of confidence in achieving particular goals or an overall optimum, and strongly consider the weight of evidence.

Evaluation of the Alternative in Relation to the Nine Evaluation Criteria and other Considerations

Table G-1: Concentration-Based Standards

Short-term effectiveness: may result in unusual circumstances of risks associated with remedial actions required to meet the standard.

Long-term effectiveness: promotes treatment by defining clean as opposed to protectiveness, achieveing clean, less emphasis on institutional control.

Reduction of toxicity, mobility, or volume: promotes treatment by requiring actions to meet defined clean levels as opposed to allowing institutional prtoective measures.

Implementability: the logistics of developing such standards and reaching agreement appears difficult. However, once the standards are developed implementability and defensibility would be improved, if the standards prove to be attainable in practice.

Cost: not considered, not allowing a balancing of the total Fund.

Compliance with ARARs: given that chemical-specific standards (ARARs) would be adopted, this alternative would insure complete compliance.

Overall protection of human health and the environment: Assuming standards are established as protective, protectiveness insured in cases similar to the generic exposure scenario. Multiple chemicals and routes may result in instances where total risk may not be protective.

Community and State Acceptance: By using uniform agreed upon standards, acceptance should be enhanced or facilitated by removing a potential point of argument.

Program Efficiency: Poorly addresses this aspect due to the inability to balance multidisciplinary concerns

Equity: using uniform standards treats individuals (rural vs. urban) equitably in terms of individual risks but not in the considerations of costs per death avoided.

Evaluation of the Alternative in Relation to the Nine Evaluation Criteria and other Considerations

Table G-2: Designated Technology (BDAT)

Short-term effectiveness: by designating remedial action without speciifc concerns related to risks associated with the implementation it is likely some increased short-term risks may result.

Long-term effectiveness: requires treatment by linking such technologies to concentration and medium and allows institutional controls in only limited circumstances.

Reduction of toxicity, mobility, or volume: requires treatment except in unique and unusual circumstances.

Implementability: with limited experience to date in remedial actions, the technology, and it's performance the logistics of developing such standards and reaching agreement appears difficult and unlikely in the short-term. However, once more information is available and the standards are developed implementability and defensibility would be greatly improved.

Cost: considerations are inherently incorporated into the process with the use of accepted proven BDAT in most circumstances. However, when applied to large, complex, or multi-contaminant situations the use of multiple BDATs may lead to excessive and potentially inefficient costs, unless provisions are in place to prevent this.

Compliance with ARARs: given that chemical-specific standards would be not be specificaly adopted and technology as opposed to concentration be the driving force, compliance with ARARs is not assured. The development of standards (i.e., % reduction) should include the existence of such chemical-specific ARARs and try to approach them but would not be required to do so. Compliance determined by the extent ARARs included in the treatment range.

Overall protection of human health and the environment: BDAT does not directly measure or meet a standard of risk-based protectiveness. Health concerns are addressed in the development or acceptance of BDAT for a particular constiuent or class of compounds but additivity and unusual exposure characteristics are not adequately addressed in the BDAT approach. Environment would be protected more in locations of limited human exposure than at present.

Community and State Acceptance: By using uniform remedial solutions acceptance should be enhanced or facilitated by removing inconsistency as a potential point of contention, but does creates other potential contentions, removing

flexibility does not allow a state or community to apply their desires (e.g., non-reliance on incineration, avoiding high cost remedies).

Program Efficiency: The use of BDAT would expedite the cleanup process by streamlining the selection process, and by utilizing technically and economically feasible technologies should promote treatment and cost efficiencies.

Equity: using uniform standards of treatment treats individuals (rural vs. urban) equitably in terms of what is done and in individual risks. By the designation of a technology the discussion of inherent value of life are circumvented but the costs of death avoided would be higher in rural locations.

Evaluation of the Alternative in Relation to the Nine Evaluation Criteria and other Considerations

Table G-3: Improved Risk Assessment

Short-term effectiveness: would improve existing risk assessment, increase confidence, improve the site characterization process by identifying sensitive parameters. Would not expedite the process only improve the quality.

Long-term effectiveness: would improve overall confidence in the results (residual cleanup levels, risk), but would have no influence over the type of technologies chosen.

Reduction of toxicity, mobility, or volume: would only influence this in circumstances where the risks of future failures are significant, and confidence in prediction is high enough to warrant additional expense than would normally be expended.

Implementability: because an existing system is augmentated, no major changes are required, therefore this alternative would have the greatest ease of implementability

Cost: while not really an integral consideration of the alternative it would likely improve cost effectiveness by allowing better comparisons between alternatives.

Compliance with ARARs: no change made from existing methodology in this regard.

Overall protection of human health and the environment: overall confidence in the measure of protectiveness would be improved in regard to human health, and to environmental effects as well if they should be included in the improvement effort.

Community and State Acceptance: by improving the risk assessment portion of the selection process and hopefully improving risk communication by utilizing a fuller definition of risk acceptance of these decisions would be improve relative to the portion that risk assessment plays in the overall acceptance.

Program Efficiency: by improving confidence in the predictions of effects, and allowing a common comparison between alternatives, and sites on the basis of level of confidence decisions would be improved and hopeully more effective actions overall would be chosen.

Equity: because this alternative will hopefully reduce and explicitly state uncertainty equity as influenced by differing uncertainties in analysis and projections will be improved. In other words there would be less influence of different degrees of uncertainty in analysis.

Evaluation of the Alternative in Relation to the Nine Evaluation Criteria and other Considerations

Table G-4: Benefit-Cost Analysis

Short-term effectiveness: depending on the complexity of the framework of analysis developed abd adopted, selection of remedial actions may actual take longer. The benefit-cost analysis would be an additional layer of analysis over that which presently exists.

Long-term effectiveness: by valuation of potential future costs due to ineffectiveness of remedial actions, long-term effective solutions would be promoted under the benefit-cost analysis approach.

Reduction of toxicity,mobility, or volume: not clear as to whether the benefit-cost analysis approach would promote treatment or improved insititutional controls.

Implementability: the potential for implementing such a program appears to be unlikely in the short-term. Significant techniques of analysis would have to be developed and adopted.

Cost: the goal of benefit-cost analysis is to promote the most economically efficient solutions to a given problem. For this reason by definition the benefit-cost analysis alternative would promote cost effectiveness.

Compliance with ARARs: there is no provision or requirement in the benefit-cost analysis approach for the meeting or exceedance of ARARs. The appropriateness of each ARAR would be determined based on economics and feasibility on a case by case basis.

Overall protection of human health and the environment: protection of the human health and the environment is determined on the basis of economics and does not on rely on some quantified measure of protectiveness as presently exists. By implication this alternative would have relative levels of protectiveness, related to the value of the resources being protected.

Community and State Acceptance: acceptance on either the state or community level is dependant on the acceptance of the framework and assumptions in the analysis. Communities and states may conceivably disagreed with the assumptions for valuation and feel as though their specific resources are being undervalued. Benefit-cost analysis by merely adding another layer of analysis does nothing to allay the concerns or mistrust of the public in the risk assessment and selection process in general.

Program Efficiency: by maximizing the efficiency of each particular decision, overall program efficiency is assured from the point of view of maximizing the effectiveness of cleanups given a finite resource base. However, while the benefit-cost analysis maximizes efficiency on an economic basis, it does not promote or insure statutory efficiency (i.e., meeting specific statutory requirements) as the goals of an efficient economic solution may be contradictory to certain statutory requirements (e.g., meeting ARARs, protectiveness).

Equity: if consistent assumptions or values for a life are to be used (if human lives, as opposed to environmental resources, dominate the analysis), larger population densities may be favored, and therefore urban areas would be favored over rural areas. Equitey would be insured only from the perspective of the assumption used in the valuation of a life. If cost per death avoided is the used, idividual risk is valued higher in an urban setting, while if a consistent value of individual risk (life) is used, the cost per death avoided would be greater in a rural setting.

TABLE G-5
RANKING/COMPARISON OF ALTERNATIVES TO EVALUATION CRITERIA

Criteria	Alternative* I	II	III	IV
Short-term Effect.	o	o	+	+
Long-term Effect.	++	+	++	-
Reduction of Mobility, volume, and/or toxicity	++	+	++	-
Implementability	++	+++	+	-
Cost	-	++	++	+++
Overall Protect.	++	++	+++	+
Compliance w/ARARs	+++	+	++	-
Community and State Acceptance	++	++	++	-
Program Efficiency	-	++	++	+++
Equity	+	+	++	+++
Completeness	--	-	++	+
Rigor of thought	-	-	++	+
Consistency	+	+	++	-
Demonstrability	+	+	++	+
Consistency**	Outcome	Outcome	Process	Process

*Alternative I: Concentation-Based Standards
Alternative II: Designated Technology (BDAT)
Alternative III: Improved Risk Assessment
Alternative IV: Benefit-Cost Analysis

**Consistency in outcome relates to outcome only (i.e., identical residual concentration or technology chosen) while consistency in process is only concerned with how the analysis is carried out with no control over outcome. However, identical characteristics should lead to identical outcomes.

NOTE: Rankings are subjective and relative, not absolute. Read text to aid in interpretation of the results.

H. FURTHER STUDY NEEDS

As stated in the previous section, additional information is needed before one can truly assess the viability of any alternative. The Superfund program can still be considered in it's infancy with limited experience. Before one can develop or select alternatives to make improvements in a given system, one must first know whether problems exist and, if they do, what are their causes. The mere identification of symptoms to a problem does not suggest a problem, but implies a diagnosis of the problem is required. While it is thought that the present application of risk assessment within Superfund has problems, the causes of the problems have not been addressed. It has been suggested that the problem is not within the science of risk assessment but in it's application.

Once the cause of problems have been properly identified then responsible alternatives can be developed remedying the problem, not just the symptom. Further study needs must be focused on both the use of risk assessment and the nature of the problem which Superfund addresses.

Risk Assessment

A thorough characterization of the present preparation and use of risk assessment within the Superfund program must be carried out to determine what elements, if any, are responsible for inconsistencies or problems. The assumption that risk assessment as the basis for establishing cleanup levels is the problem may be erroneous leading to misguided and futile efforts to make changes. A complete statistical review of past risk assessments should be made to reveal assess inconsistencies.

All risk assessments prepared to date should be statistically analyzed. Specific studies should focus on the following:

- assumptions used and the values for common parameters including exposure variables;
- the use of available scientific evidence (e.g., whether information used for specific chemicals is consistent across sites); and
- use of models (dose-response, environmental transport)

The statistical analysis should determine what is the mean and variance for each of these, and whether they're used consistently. Deviations from the norm should be quantified as a percentage of the total. The norms indicated by this analysis should be compared to what would be scientifically acceptable. If deviations from the norm are found these should be investigated further. Determinations should be

made if these deviations are justified, related to site-specific conditions, or made at the discretion of the preparer or reviewer.

Uncertainty within past risk assesment should be quantified and compared as best possible. The comparison of uncertainty between alternatives can give a picture of how decisons have been made and whether they are efficient.

It should be noted that the Toxics Integration Branch within OERR is supervising a computer-based project reviewing risk assessments for quality control purposes. A computer program, using a generic exposure scenario and assumptions, would calculate risk levels with site-specific information inputs. The result would then be compared to the results reported in the prepared risk assessment. Deviation between the two would then be calculated. If the deviation between the two is within an acceptable range, the risk assessment is considered acceptable. Any deviation greater than a predescribed amount flags that risk assessment for further review. While this project is targetted for quality control it may also be a useful tool to gather information on the risk assessments being performed.

A study should be made comparing risk assessments prepared and the final cleanup levels between Principal Responsible Parties (PRPs)-led and Fund-led sites . A comparison of the two would determine if in fact the risk assessments are subject to manipulation (or greater detail, sophistication), especially on the part of PRPs. The comparison of the two should focus on: the assumptions used, values used (for exposure parameters), final level of protectiveness, and the degree of uncertainty. A comparison of uncertainty may be quite interesting as this is the area which can be most influenced by the professional discretion of the preparer (through the increased use of models and data). Models and data different than accepted values may have been justified which yield more favorable results for the PRP.

Site Characterization/Categorization

The BDAT approach is contingent on the ability to categorize Superfund sites into a workable number of categories. A statistical review must be carried out on many of the parameters of site characteristics to determine if categorization of the sites into a workable number of categories (about 10-12) is possible. Parameters which should be addressed include site geology, hydrology, and soils. The percentage of time that site-specific conditions justify a deviation from a norm needs to be estimated. These exceptions should also be categorized.

Resource Valuation

To promote or support a benefit-cost analysis approach further studies would be required in the area of resource valuation and benefits assessment. There are techniques available and effort should be made to apply them to the specific problems of Superfund sites. Without the additional information from these studies the benefit-cost analysis alternative would not be feasible.

I. CONCLUSIONS AND RECOMMENDATIONS

The remedial selection process as defined by the RI/FS process is believed to be adequate in scope. This opinion is universal across all interests. Therefore the problems associated with the Superfund program are not related to design. The problems are due primarily to inconsistent application. Even the major critics of the Superfund program have presented case studies believed to be exemplary which demonstrate or achieve the goals established in the Superfund legislation (OTA, 1988, EDF et al, 1988).

As there is agreement that the design of the program is sufficient, one must seriously question the need for change or redirection. The use of risk assessment is presently used as the basis for establishing cleanup levels, and has shown to sufficiently achieve the goals of the program in several examples. The abandonment of risk assessment as the basis of these decisions would be short-sighted. Any new alternative foundation for these decisions would require time for development and implementation. Likewise any alternative would have its own limitations and problems.

In other words, the process of remedial action selection is believed to satisfy the requirements of completeness of conception. Further improvements are still possible in the areas of rigor of thought and consistency of application. Efforts should concentrate on developing more rigorous or formal procedures without sacrificing completeness. While formal procedures are required the ability to incorporate information from broad sources (empirical to theoretical) should be maintained as best as possible.

The conclusions are supported by the fact that the Superfund program can be considered to still be in its infancy with limited experience. As experience is gained new knowledge is obtained which can then lead to focusing or expanding the program. Knowledge gained from information about shortcomings should be used to improve the program not abandon it. Abandonment should only be chosen if experience shows the exisiting process to be inadequate. This is certainly not the case for the RI/FS.

While uncertainties in risk assessment may be great and vast improvements are neeeded there appears to be no useful alternative to its use.(Rodricks et al, 1987) No alternative to risk assessment evaluated appears to yield improvements when used alone, and in actuality would likely be inferior. This is not to say that improvements are not possible in the Superfund program. In fact, the optimum is a combination of that described here. Each of the alternatives represents an integral component of the overall process. Therefore improvements to these components would yield overall improvement.

While the other elements should contribute to the overall decision and design, the establishment of cleanup levels at remedial actions should still be primarily based on (improved) risk assessment. No other alternative appears adequate for this purpose. However, essential to the future success of Superfund is the ability to incorporate environmental and natural resource impacts into the risk assessment methodology. At present the remedy selection process is dictated by human effects and do not adequately address environmental impacts.

The decision process should continue to be an aggregate with all elements playing a vital role. Future emphasis should be placed on improving all elements contributing to the decision of adequate cleanup levels and incorporate many elements of each alternative evaluated. While the basis for setting cleanup levels will continue to be primarily based on risk they are not likely to be as rigid as at present. With an expanded definition of risk decision can be made trading off with other elements. For example, settling for a lower confidence in the risk estimate because of costs of achieving higher levels may be prohibitive.

The EPA is presently taking the proper course of action by consolidating and evaluating information on prepared risk assessments, technology, and site characteristics. This information would be used to improve guidance and consistency throughout the program, and thereby improve the quality of the decisions for remedial action design. Continuation and accelerating these activities should prove advantageous.

Short-term Recommendations

Decisions are presently required to balance the goals and objectives (nine evaluation criteria) without the aid of any guidance as to their relative weights. The most urgent short-term goal of the agency is to develop some agreement as to the relative importance of each criteria and a formal method or guidance in balancing them.

As many of the alternatives to the use of risk assessment in establishing cleanup levels at Superfund Remediaiton actions do not appear to be feasible in the near term, risk assessment shall continue to be relied upon. Immediate improvements in the remedy selection and establishment of cleanup levels can be made through improvements in the risk assessment.

The EPA should continue their efforts in defining ecological risks. Efforts in this area should be expanded and formal methods developed. Results from these efforts should be incorporated into the decision process as soon as possible.

Many of the elements described in the improved risk assessment alternative should be implemented as soon as possible, such as the greater use of available information and models, use of distributive models and estimates, and more formal quantitative sensitivity and uncertainty analysis.

Sensitivity analysis and value of information would improve the focus of the RI/FS process at all points. Proper application of these tools would eliminate the criticism of Superfund program management that too much information is sought before a decision is made, and that information obtained is of questionable quality. This image of decision paralysis would be shed. Techniques of multi-attribute optimization can be used in comparison of alternatives in applying the nine evaluation criteria.

The US EPA/OAQPS has previously funded a program at Argonne National Laboratory to investigate the feasibility and the value of incorporating decison analysis techniques in the regulatory process for toxic air pollutants. An example of efforts to incorporate these techniques into a regulatory decision are the in the EPAs review of the primary NAAQS for lead (Zarragoza,1987). This effort should be reviewed and efforts made to apply the same techniques and EPAs experiences to the Superfund program.

Preliminary phases of study should be initiated for those identified in Section H. to determine the feasibility of the alternatives discussed. Each of the alternatives is contingent on additional information.

Long-term Recommendations

Long-term recommendations are contingent on the results of initial studies and the changing goals of the EPA and the nation as a whole. Results of the initial studies would indicate the feasiblity of each alternative. Likewise which of the nine evaluation criteria or goals of the program are emphasized in the future would influence the desirability of any particular alternative.

It is the opinion of the author that no one alternative be selected and emphasized at the expense of the others. Superfund is unique in that it is a conglomeration of all previously existing programs and standards. The breadth of it's concern is unequalled in previous programs. Because of the resulting complexity of meshing standards and programs no one alternative would be sufficient to handle the objectives and intent of the legislation. Site-specificity is usually used as the explanation for having flexibility within the RI/FS process. However, a more far reaching reason for the flexibility is in the balancing of different

programs and standards. Therefore the interactive nature of the RI/FS should be maintained, improvements if necessary should be accomplished through improvements to each of the individual elements.

APPENDIX A

Simplified Representation of Cumulative Probability Distributions.

This will be a very simplified demonstration of possible uses and interpretations of probability density functions and cumulative probability distributions in risk estimation and management decisions. This demonstration will consist of a simplified example designed to express a point. For greater detail on cumulative probability and how they are derived one can consult the literature (e.g., Granger et al, 1987, Crawford-Brown, Cothern, 1987) This example is designed to show how greater information is contained in such a presentation of risk.

For the purposes of this example Figure 1 and 2 represent the risks associated with two chemicals A and B. Figure 1 represents the probability density of predicted effects, and Figure 2 represents the cumulative probability. For simplicity sake it can be assumed that 100 studies have been carried out on each chemical and the results of each study represent a single percentage point. The graph is read by comparing the value with the overall confidence. That is one is 100% confident that the number of deaths expected would be X_o or less. Likewise the best estimate, associated with 50% confidence, suggests that the number of deaths for compound A would be X_2 or less, and that no deaths would be expected for compound B.

Based on conventional EPA risk assessment methodology there would be little difference between the two compounds. The range of risks would be between zero and X_o number of deaths expected and the 95% confidence level (plausible upper bound estimate) would be equal at X_1. Conventional risk assessment does not use all available information and would stop there.

The cumulative probability distribution yields greater information on differences between the two. There is significant evidence that compound A has severe health effects while for compound B, while evidence exists, it is much more limited. In fact the majority of evidence tends to suggest that there is little or no anticipated health consequences from compound B. This may be analogous to compound A being a known carcinogen and compound B being classified as a possible carcinogen.

The differences between the two, utilizing the Figure 2 would have implications on decisions for remedial actions. Given the differences in evidence between the two compounds it would be irresponsible to treat them equally in risk management decisions, either in apportionment in risk

FIGURE 1

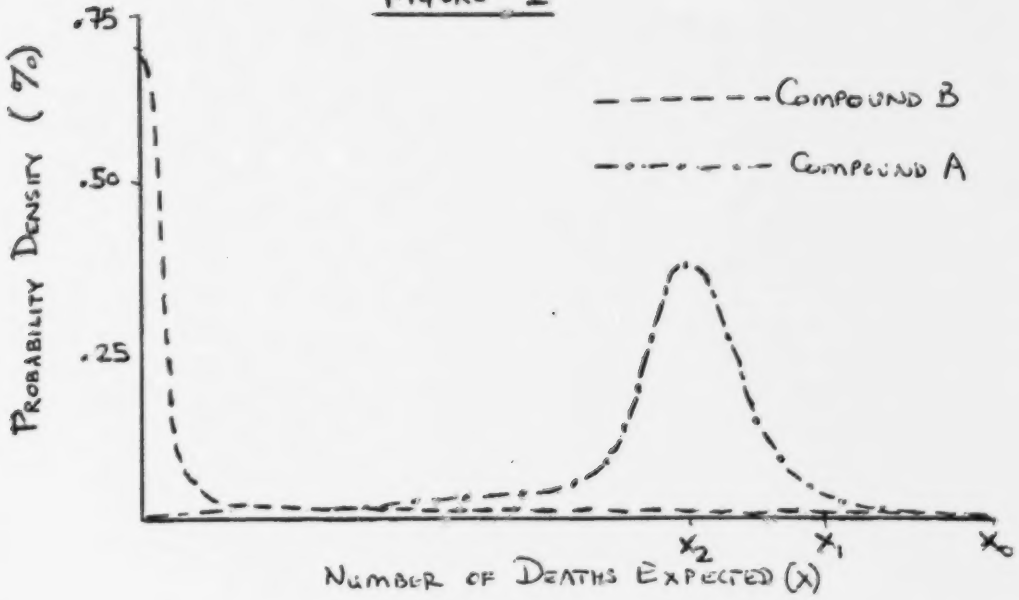
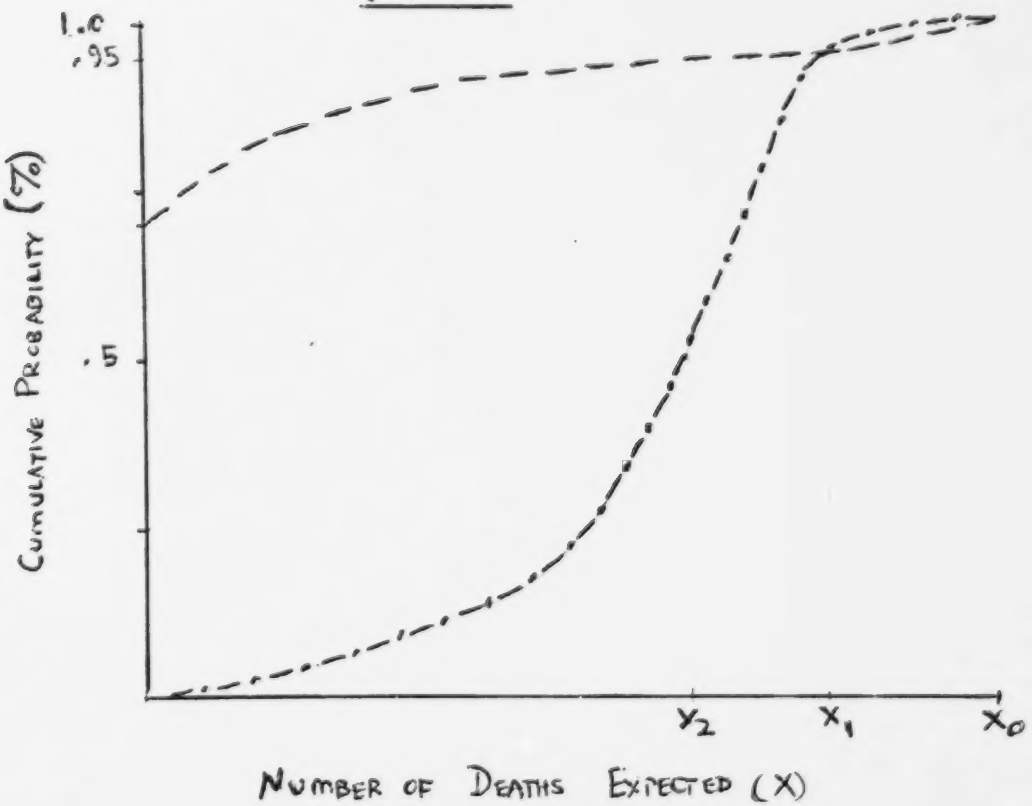


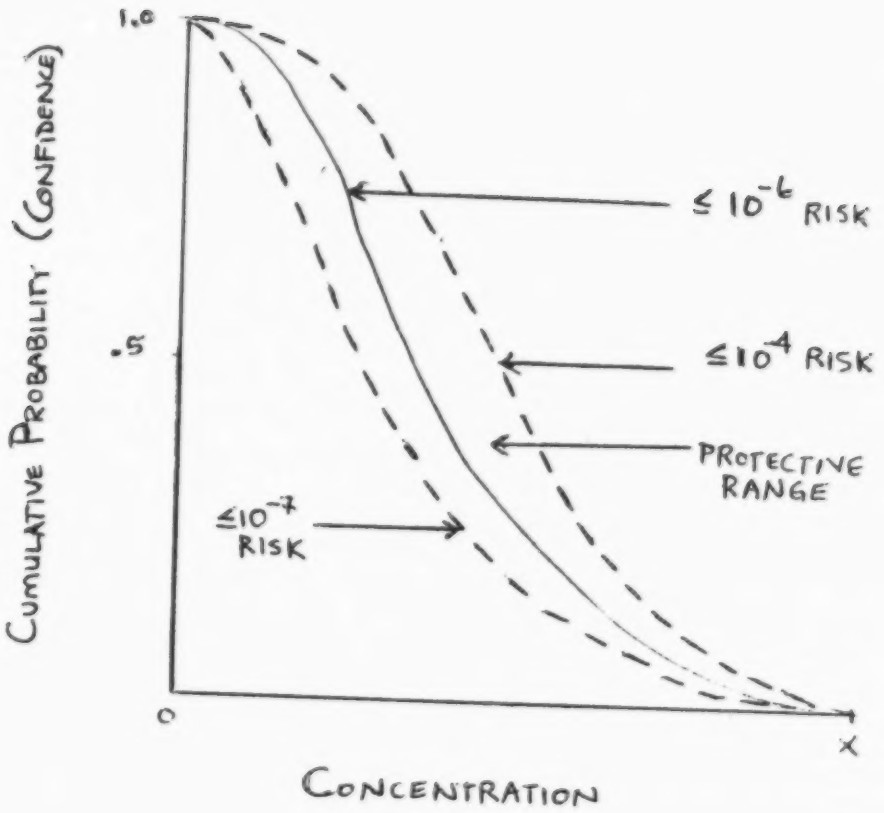
FIGURE 2



assessment or in designing and selecting remedial actions. Based on the differences between the two compounds actions or emphasis should be directed to compound A.

Conventional risk assessment carried out by the EPA does not make or utilize this distinction. Typically the negative study results (as in compound B) would typically be discounted in the existing databases. This is a very valuable tool which should be utilized to a greater extent in the future. While this simplified example addresses the estimated number of deaths expected at a given exposure the same approach can be utilized for many other applications. For example Figure 3 shows the relationship of the risk range to concentration for a given compound (either intake or extrapolated environmental residual concentration).

FIGURE 3


1.0
.5
0
X
CUMULATIVE PROBABILITY (CONFIDENCE)
CONCENTRATION
≤ 10⁻⁶ RISK
≤10⁻⁴ RISK
PROTECTIVE RANGE
≤10⁻⁷ RISK

APPENDIX B: LIST OF INTERVIEWEES

CATEGORY 1

U.S. Environmental Protection Agency

OSWER:	
David Bennett	Toxics Intergration Branch
Bruce Means	
Sandra Lee	
Arthur Weisman	Policy Analysis Branch
Steve Smagin	Remedial Planning and Response Branch
Hal Snyder	Design and Construction Man. Branch
William Hanson	Site Policy & Guidance Branch
Larry Reed	Hazardous Site Evaluation Division
Art Day	RCRA, Land Disposal Branch
Alec McBride	Technical Assessment Branch, OSW
Sherry Sterling	RCRA, Compliance
John Cunningham	
Larry Zarragoza	Office of Cross-Media Analysis
Kevin Garrahan	Office of Research and Development
Thomas Evans	
Craig Zamuda	Office of Policy and Program Evaluation
Skip Luken	OPPE, Economics Analysis Branch
Terry Dinan	
Steve Clark	Office of Drinking Water
Jennifer Orme	
Frank Gostomski	Office of Water
Karen Blanchard	Office of Air

CATEGORY 2

Mark Reiter	Senate Committee on the Environment and Public Works
Kirsten Oldenberg	Office of Technology Assessment
Dave Gushy	Environment and Natural Resources Div., Congressional Research Service
Mark Reisch	"
James Solyst	National Governors Association

CATEGORY 3

Brad Brockbank	Consevation Foundation
Clarence E. Davies	"
Lois Epstein	Environmental Defense Fund
Ellen Silbergeld	"
Glen Anderson	Environmental Law Institute
Adam Finkel	Resources for the Future

CATEGORY 4

Tim O'Leary	Chemical Manufacturers Association
Geraldine Cox	
Janet Matey	
Paul Price	American Petroleum Institute